Rowan University

Rowan Digital Works

7-29-2015

Level set segmentation using non-negative matrix factorization with application to brain MRI

Dimah Dera

LEVEL SET SEGMENTATION USING NON-NEGATIVE MATRIX FACTORIZATION WITH APPLICATION TO BRAIN MRI

by

Dimah Dera

A Thesis

Submitted to the
Department of Electrical & Computer Engineering
College of Engineering
In partial fulfillment of the requirement
For the degree of
Master of Science in Electrical and Computer Engineering
at
Rowan University
May 15, 2015

Thesis Chair: Nidhal Bouaynaya, Ph.D.

Dedications

I would like to dedicate this work to my mother, my father, my aunt and my husband.

Acknowledgments

Abstract

Dimah Dera
LEVEL SET SEGMENTATION USING NON-NEGATIVE MATRIX
FACTORIZATION WITH APPLICATION TO BRAIN MRI
2014-2015
Nidhal Bouaynaya, Ph.D.
Master of Science in Electrical & Computer Engineering

We address the problem of image segmentation using a new deformable model based on the level set method (LSM) and non-negative matrix factorization (NMF). We describe the use of NMF to reduce the dimension of large images from thousands of pixels to a handful of "metapixels" or regions. In addition, the exact number of regions is discovered using the nuclear norm of the NMF factors. The proposed NMF-LSM characterizes the histogram of the image, calculated over the image blocks, as nonnegative combinations of basic histograms computed using NMF ($V \approx W H$). The matrix W represents the histograms of the image regions, whereas the matrix H provides the spatial clustering of the regions. NMF-LSM takes into account the bias field present particularly in medical images. We define two local clustering criteria in terms of the NMF factors. The first criterion defines a local intensity clustering property based on the matrix W by computing the average intensity and standard deviation of every region. The second criterion defines a local spatial clustering using the matrix H. The local clustering is then summed over all regions to give a global criterion of image segmentation. In LSM, these criteria define an energy minimized w.r.t. LSFs and the bias field to achieve the segmentation. The proposed method is validated on synthetic binary and gray-scale images, and then applied to real brain MRI images. NMF-LSM provides a general approach for robust region discovery and segmentation in heterogeneous images.

Table of Contents

Table of Contents (continued)

List of Figures

List of Tables

Chapter 1

Introduction

In this section, we will motivate and state the problem of medical image segmentation, review the state-of-the-art approaches applied in this field and shed the light on the main contributions of our thesis work.

1.1   Motivation, Problem Statement and Background

Medical image segmentation is one of the most important and complex tasks in medical image analysis and is often the first and the most critical step in many clinical applications, such as surgical planning and image-guided interventions. For instance, in brain MRI analysis, we need to visualize and measure the brain anatomical structures, detect the changes in the brain and delineate the pathological regions. Segmentation of brain MRI images into specific tissue types requires assigning to each element or pixel in the image a tissue label, where the labels are defined in advance. For normal brain, image pixels are typically segmented into three main tissue types: white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF), while in the case of brain tumors, such Glioblastoma, there are additional structures that include the tumor, edema (swelling) and necroses (dead cells).

There are three main challenges with brain MRI segmentation: i) The (normal and abnormal) brain anatomical structures have complex morphologies and boundaries; ii) The distinct regions of the brain MRI are not homogeneous but present an intensity inhomogeneity, or bias field. The bias field arises from the spatial inhomogeneity of the magnetic field, the variations in the sensitivity of the reception coil and the interaction between the magnetic field and the human body [37]; and iii) Distinct anatomical structures may have

close average intensity values, e.g., gray-matter and necrosis. These challenges make classical segmentation techniques, such as thresholding [36], edge detection [16], [17], region growing [34], [43], classification [15], [42], [44], and clustering [1], [9], [14] ineffective at accurate delineation of complex boundaries.

Deformable models are ones of the most powerful and advanced methods for image segmentation. The basic idea is to evolve a curve in the image domain around the object or the region of interest until it locks onto the boundaries of the object. The deformable model segmentation problem is formulated and solved using calculus of variations and partial differential equations (PDEs). Deformable models can be classified into two groups: *snakes* or *active contour* models [6], [12], [20], [21] and *level set* methods [31], [32], [35]. In the snake or active contour model, the contour is represented in a parameterized form by a set of points that are propagated under the influence of an internal energy and an external energy. The internal energy defines the shape of the contour and imposes smoothness and relevant geometrical constraints on the curve. The external energy is computed from the image and attracts the contour towards objects boundaries and other desired features in the image. However, the major drawbacks of the active contour model are its sensitivity to the initial conditions and the difficulties associated with the topological changes for the merging and splitting of the evolving curve. These difficulties actually lie in the parametric representation of the contour. For instance, when the contour merges and splits to fit the objects boundaries in the image, one has to keep track of which points are in which contour and what their order is. The level set approach proposes a geometric (rather than parametric) representation of the contour. Specifically, the contour is represented as the zero level of a higher dimensional function, referred to as the level set function. Instead of tracking

a curve through time, the level set method evolves a curve by updating the level set function at fixed coordinates through time. In particular, since the level set does not have any contour points, the merging and splitting of the curve is done automatically and no new contours need to be defined or removed. The internal and external energies, in the level set approach, are defined in a similar manner as in the active contour method. The power of the level set and active contour methods, referred to as deformable models, stems from their continuous formulation, which can achieve pixel-level accuracy, a highly desirable property in medical image segmentation.

A good body of the work has been done to develop an accurate and robust external energy, also called *the data term*, that can move the curve and accurately fit the regions boundaries in the image. One of the earliest level set approaches is the Mumford Shah (MS) model [39]. It assumes that the image is piecewise smooth in the areas of objects and backgrounds. However, it is difficult to apply the gradient descent method to solve the MS model because of the non-differentiability with respect to the image boundary. Chan and Vese [7] simplified the MS model using a variational level-set formulation. The Chan-Vese model is based on the assumption that the intensity within each region is homogenous or roughly constant. The image is thus approximated by a constant inside every region. This model, however, is only effective for piecewise constant images and it does not handle intensity inhomogeneity within regions. A major shortcoming of the MS and Chan-Vese models is their assumption of intensity homogeneity within each region of the image.

Recently, local intensity information has been incorporated into the level set methods to effectively handle intensity inhomogeneity. In [26], Li *et al.* defined a region-scalable fitting (RSF) energy functional as the external energy term of the level set by using a kernel

function with a scale parameter, which allowed the use of local intensity information in image regions at a controllable scale, and two fitting functions that locally approximate the image intensities on the two sides of the contour. The RSF model simultaneously estimates the local intensity mean of each region with the level set function through an iterative procedure. This local definition of regions statistics was the only way to handle the intensity inhomogeneity in the RSF model. In [41], Wang *et al.* proposed a local Gaussian distribution fitting (LGDF) energy with a level set function by also using a kernel function and local means and variances as variables, that were also simultaneously derived with the level set function in an iterative procedure. Both the RSF and LGDF models relied on estimating the local statistics of the image regions through the level set formulation to handle intensity inhomogeneity without introducing the bias field as a separate variable to correct for the intensity inhomogeneity in the original image. More recent applications of the level set approach, which also took into account the intensity inhomogeneity, defined a local clustering criterion for the image intensities in a neighborhood of each pixel. This local clustering was then integrated to give a global criterion of image segmentation, and served as the external energy term of the level set formulation. These methods used localized clustering (Localized-LSM) [25], and statistical characteristics of local intensities (Improved LGDF-LSM) [10]. Obviously, the performance of the level set approach depends on the local clustering criterion used. These two methods added the bias field as a separate variable estimated through the variational principle of the level set formulation in order to correct for the intensity inhomogeneity in the original image. Multiplying the local average intensity of each region inside the neighborhood by the bias field variable gives us different intensity values in the same region and thus handles the intensity inhomogeneity in every

4

region. Moreover, each local clustering criterion has its own parameters that have to be simultaneously estimated along with the level set function and the bias field. For instance, in the localized level set clustering criterion (Localized-LSM) in [25], the local intensities in every region have to be estimated along the bias field and the level set functions. Similarly, the statistical approach (Improved LGDF-LSM) in [10] involved simultaneous and iterative estimation of the mean and variance and other parameters of the local density approximation.

All previously mentioned approaches involved simultaneous and iterative estimation of a number of model parameters in addition to the bias field and the level set function, which is the main parameter to be estimated. Given the high-dimensionality and non-convexity of the variational optimization problem, all additional model parameters are estimated in an iterative procedure that does not guarantee convergence or optimality of the results. Hence, one of the drawbacks of these state-of the-art approaches is the number of model parameters that are introduced, and have to be simultaneously estimated, which decreases the estimation accuracy of the main segmentation parameters, namely the level set functions. Moreover, all recent level set approaches are built based only on the first and second order statistical features: the mean and standard deviation intensities of the pixel values. Moreover, these features do not incorporate any information on the spatial distribution of the pixel values which may greatly improve the segmentation.

1.2   Research Contributions

This thesis contributes to the field of (medical) image segmentation by introducing a new deformable model that is able to delineate complex boundaries of image regions by relying

on the image density information (histogram) rather than the absolute pixel intensity values and correcting for the intensity inhomogeneity without introducing additional model parameters to be estimated simultaneously with the level set functions. The proposed segmentation framework has four advantages compared to the state-of-the-art: i) less sensitive to the model parameters, ii) more robust to noise in the image, iii) less sensitive to the initial contour, and iv) has a higher convergence rate. Specific contributions of this work include:

- Building the data matrix using the histograms of the image blocks rather than relying on the absolute pixel intensity values. This characterization makes the subsequent algorithm more robust to noise in the image.

- Elucidating how the non-negative matrix factorization (NMF) of the data matrix can cluster the image into distinct regions, and how relevant image structures can be extracted from the NMF factors.

- Deriving a measure based on the nuclear norm of the NMF factors to estimate the number of distinct regions in the image.

- Introducing two new external energy or data terms derived from the two factors of the NMF. In particular, introducing a new spatial term, which increases the resolution of the proposed algorithm by increasing its ability to discriminate between distinct regions with close average intensity values.

- Proposing a new NMF-based level set method with the bias field correction to take into account intensity inhomogeneity. The proposed model does not require the estimation of spurious model parameters in addition to the bias field and the level set

6

functions.

- Incorporating the statistics of noise in the data by using probabilistic non-negative matrix factorization (PNMF) given in [4], which assumes that the data matrix is corrupted by additive white Gaussian noise.

1.3 Organization

This thesis is organized as follows.

In Chapter 2, we provide a literature review of the state-of-the-art level set models, describing their mathematical formulation and model parameters. It is crucial to understand the mathematical and theoretical assumptions of the previous work in order to grasp the novelty of this thesis.

In Chapter 3, we review the mathematical and theoretical formulation of the non-negative matrix factorization (NMF) and its variants, including probabilistic NMF (PNMF). PNMF assumes that the data matrix is not deterministic but corrupted by additive white Gaussian noise. We subsequently explain how NMF can be used to discover the image regions and cluster them.

In Chapter 4, we review the mathematical and theoretical formulation of the level set method (LSM), including the level set membership functions that partition the image domain into disjoint, non-overlapping regions.

In Chapter 5, we introduce the proposed PNMF-LSM approach. We describe how the positive factors of the PNMF discover and cluster the image domain into distinct regions. We introduce two external energy terms that will drive the contour to the regions bound-

aries. We take into account the bias field and carry out the segmentation by minimize the total energy functional with respect to the level set functions.

In Chapter 6, we provide and discuss the simulation results by evaluating the performance of the proposed PNMF-LSM method as compared to two other state-of-the-art level set methods, the localized level set model (localized-LSM) [25] and the improved LGDF level set model (improved LGDF-LSM) [10]. We study the robustness of the models to the model parameters, initial conditions and noise introduced in the image. We also discuss the convergence time of the models.

In Chapter 7, we apply the proposed PNMF-LSM method to real brain MRI images, with and without tumor, to delineate the complex structures of the brain: gray matter, white matter, cerebrospinal fluid (CSF), edema (swelling), tumor and necroses (dead brain cells). We also show the robustness of our method to blurring by Gaussian kernels and to salt and pepper noise.

Finally in Chapter 8, we provide a brief conclusion that summarize this work.

Chapter 2

Literature Review

In this chapter, we will review the state-of-the-art in level set methods that build different data terms (or external energies) in the level set framework for image segmentation.

2.1 Mumford-Shah Model [39]

Let Ω be the image domain, and $I : \Omega \to \mathbb{R}$ be a gray-value image. The goal of the segmentation is to find a contour C, which separates the image domain Ω into disjoint regions $\Omega_1, \cdots, \Omega_k$, and a piecewise smooth function u that approximates the image I and is smooth inside each region Ω_i. This is formulated as the minimization of the following Mumford-Shah functional:

$$\mathcal{F}^{MS}(u, C) = \int_\Omega (I - u)^2 d\boldsymbol{x} + \mu \int_{\Omega \backslash C} |\nabla u|^2 d\boldsymbol{x} + \nu |C|, \qquad (2.1)$$

where $|C|$ is the length of the contour C. In the right hand side, the first term is the external energy term, which drives u to be close to the image I, and the second term is the internal energy, which imposes smoothness on u within the regions separated by the contour C. The third term regularizes the contour. The MS model is very general and does not assume a specific form for the approximating function u. It also assumes that the objects to be segmented are homogeneous.

2.2 Chan and Vese Model [7]

Chan and Vese simplified the Mumford-Shah model by assuming that the approximating function u is a piecewise constant:

$$
\mathcal{F}^{CV}(\phi, c_1, c_2) = \int_{\Omega} |I(\boldsymbol{x}) - c_1|^2 H(\phi) d\boldsymbol{x}
$$
$$
+ \int_{\Omega} |I(\boldsymbol{x}) - c_2|^2 (1 - H(\phi)) d\boldsymbol{x} + \nu \int_{\Omega} |\nabla H(\phi)| d\boldsymbol{x}, \tag{2.2}
$$

where H is the Heaviside function, and ϕ is a level set function, whose zero level contour C partitions the image domain Ω into two disjoint regions $\Omega_1 = \{\boldsymbol{x} : \phi(\boldsymbol{x}) > 0\}$ and $\Omega_2 = \{\boldsymbol{x} : \phi(\boldsymbol{x}) < 0\}$. Equation (2.2) is a piecewise constant model, as it assumes that the image I can be approximated by constants c_i in region Ω_i. In the case of more than two regions, two or more level set functions can be used to represent the regions $\Omega_1, \cdots \Omega_k$.

2.3 Localized-LSM Model [25]

In [25], Li *et al.* proposed a variational level set method that deals with intensity inhomogeneity by considering the following model for the observed image I:

$$
I = b * J + n, \tag{2.3}
$$

where b is the bias field, J is the true image and n is the additive noise. This approach has two assumptions: a) the bias field is assumed to be slowly varying, and b) the true image J is approximated by a constant inside each region: $J(x) \approx c_i$ for $x \in \Omega_i$. Consider the neighborhood around pixel y, $O_y = \{x : \|x - y\| \leq \rho\}$, then $b(y) \approx b(x)$ inside the

neighborhood O_y. The energy function is then formulated as the following [25]:

$$\mathcal{F}(\phi, \mathbf{b}, \mathbf{c})$$
$$= \int \left(\sum_{i=1}^{N} \int K(y-x)(I(x) - b(y)c_i)^2 M_i(\phi(x)) d\boldsymbol{x} \right) d\boldsymbol{y}, \qquad (2.4)$$

where $K(y-x)$ is a non-negative weighting function that defines the neighborhood O_y, $M_i(\phi(x))$ is the membership function that represents each region using the Heaviside function, (for two regions $M_1(\phi) = H(\phi)$, and $M_2(\phi) = 1 - H(\phi)$). In the localized-LSM model, the intensity means $c_1, \cdots, c_k$ of each region are estimated iteratively along with the level set function ϕ and the bias field $\mathbf{b}$ using the variational principle of the level set framework.

2.4   Improved LGDF-LSM Model [10]

In the Improved LGDF-LSM model [10], Chen *et al.* characterize the local distribution of the intensities in the neighborhood O_x using a local Gaussian distribution. The segmentation is then achieved by maximizing the a posteriori probability. They used the log transform of the same image model in Li's method $\widetilde{I} = log(I) = log(J) + log(b)$ so that the bias becomes an additive factor rather than a multiplicative factor.

Let $p(x \in \Omega_i \cap O_x | \widetilde{I}(x))$ be the a posteriori probability of the subregions $\Omega_i \cap O_x$ given the log transform of the observed image. Using Bayes' rule $p(x \in \Omega_i \cap O_x | \widetilde{I}(x)) \propto p(\widetilde{I}(x) | x \in \Omega_i \cap O_x) \, p(x \in \Omega_i \cap O_x)$. Assuming that the prior probabilities of all partitions are equal, and the pixels within each region are independent, the MAP estimate can be achieved by finding the maximum of $\prod_{i=1}^{N} \prod_{x \in \Omega_i \cap O_x} p_{i,y}(\widetilde{I}(x))$. It can be shown that the MAP formulation can be converted to the minimization of the following energy functional

11

in the level set framework:

$$\mathcal{F}(\phi, \mathbf{b}, \mathbf{c}, \sigma^2) =$$

$$\int \sum_{i=1}^{N} \int -K(y - x) \log p_{i,y}(\widetilde{J}(x) - \widetilde{b}(y)) M_i(\phi(x)) d\boldsymbol{x} d\boldsymbol{y}, \tag{2.5}$$

where $p_{i,y}(\widetilde{J}(x) - \widetilde{b}(y))$ is modeled by a Gaussian distribution. In the improved LGDF model the intensity means $\{c_i\}_{i=1}^{k}$ and variances $\{\sigma_i^2\}_{i=1}^{k}$ of each region are simultaneously and iteratively estimated with the level set function ϕ, and the bias field $\mathbf{b}$ using the variational principle of the level set framework.

Chapter 3

Non-Negative Matrix Factorization

In this chapter, we review the theoretical and mathematical formulation of the non-negative matrix factorization and some of its variants. Then, we shed the light on the NMF as a clustering technique.

3.1 NMF and Its Variants

3.1.1 Standard-NMF. Non-negative matrix factorization (NMF) is a matrix decomposition approach which decomposes a non-negative matrix into two low-rank non-negative matrices. It was introduced as a dimensionality reduction method for pattern analysis [24]. When a set of observations is given in a matrix with nonnegative elements, NMF seeks to find a lower rank approximation of the data matrix, where the factors that give the lower rank approximation are also non-negative. The non-negativity constraint is required in some applications in order to obtain physically meaningful results and interpretations. Mathematically, the problem is formulated as follows: Given a non-negative matrix $V \in R^{n \times m}$, NMF provides two non-negative matrices $W \in R^{n \times k}$ and $H \in R^{k \times m}$ such that $V \approx WH$. The optimal factors minimize the squared error and are the solutions to the following constrained optimization problem,

$$\min_{W,H} f(W, H) = \|V - WH\|_F^2, \text{ subject to } W, H \geq 0, \tag{3.1}$$

where $\|.\|_F$ denotes the Frobenius norm and f is the squared Euclidean distance function between V and WH. Observe that the cost function f is convex with respect to one of the variables W or H, but not both. Alternating minimization of such a cost leads to the

Alternating Least squares (ALS) algorithm [22], which can be described as follows:

1. Initialize W randomly or by using any a priori knowledge.

2. Estimate H as $H = (W^T W)^- W^T V$ with fixed W.

3. Set all negative elements of H to zero or some small positive value.

4. Estimate W as $W = V H^T (H H^T)^-$ with fixed H.

5. Set all negative elements of W to zero or some small positive value.

In this algorithm, A^- denotes the MoorePenrose inverse of A. The ALS algorithm has been used extensively in the literature [2], [18]. However, it is not guaranteed to converge to a global minimum nor even a stationary point. Moreover, it is often not sufficiently accurate, and it can be slow when the factor matrices are ill-conditioned or when the columns of these matrices are co-linear. Furthermore, the complexity of the ALS algorithm can be high for large-scale problems as it involves inverting a large matrix [4]

Lee and Seung [23] proposed a multiplicative update rule, which is proven to converge to a stationary point, and does not suffer from the ALS drawbacks. The multiplicative update rule of the Lee and Seung's algorithm is shown in Eq. (3.2) as a special case of a class of update rules, which converge towards a stationary point of the NMF problem.

$$
\begin{cases}
H_{ij} \leftarrow H_{ij} \frac{(W^T V)_{ij}}{(W^T W H)_{ij}} \\
\\
W_{ij} \leftarrow W_{ij} \frac{(V H^T)_{ij}}{(W H H^T)_{ij}}.
\end{cases}
\tag{3.2}
$$

Iteration of these update rules converges to a local maximum of the objective function:

$$
F = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[V_{ij} \log(WH)_{ij} - (WH)_{ij} \right].
\tag{3.3}
$$

The update rules preserve the non-negativity of W and H and also constrain the columns of W to sum to unity. This sum constraint is a convenient way of eliminating the degeneracy associated with the invariance of WH under the transformation $W \rightarrow W\Lambda$, $H \rightarrow \Lambda^{-1}H$, where Λ is a diagonal matrix.

3.1.2 Sparse-NMF. Sparsity is a popular regularization principle in statistical modeling [38], and has been used to reduce the non-uniqueness of solutions and also to enhance the interpretability of the NMF results. The sparse-NMF proposed in [29] imposes sparsity on the factor matrix H by constraining the l_1-norm of its columns and imposes a unity-norm on the columns of W to ensure the uniqueness:

$$\min_{W,H} f(W, H) = \|V - WH\|_F^2 + \lambda \sum_{i=1}^{n} \|h_i\|_1 \tag{3.4}$$

subject to $W, H \geq 0, \|w_i\|_2^2 = 1, i = 1, ..., k.$

The optimization problem in Eq. (3.4) is solved in [29] using non-negative quadratic programming (NNQP).

For a more comprehensive overview of the different variants of NMF, including Versatile sparse matrix factorization and Kernel-NMF, the reader is referred to [29].

3.1.3 Probabilistic Non-Negative Matrix Factorization PNMF [4]. In [4], It is assumed that the data, represented by the non-negative matrix V, is corrupted by additive white Gaussian noise, and follows the following conditional distribution,

$$p(V|W, H, \sigma^2) = \prod_{i=1}^{N} \prod_{j=1}^{M} \mathcal{N}(V_{ij}|\mathbf{u}_i^T \mathbf{h}_j, \sigma^2), \tag{3.5}$$

where $\mathcal{N}(.|\mu, \sigma^2)$ is the probability density function of the Gaussian distribution with mean μ and standard deviation σ, $\mathbf{u}_i$, and $\mathbf{h}_j$ denote, respectively, the ith row of the matrix W and the jth column of the matrix H. Zero mean Gaussian priors are with standard deviations σ_W and σ_H, respectively, imposed on $\mathbf{u}_i$ and $\mathbf{h}_j$ to control the model parameters. W, and H are estimated using MAP criterion by minimizing the following function:

$$f(W, H) = \|V - WH\|_F^2 + \alpha\|W\|_F^2 + \beta\|H\|_F^2, \text{ subject to } W, H \geq 0, \qquad (3.6)$$

where the parameters α and β depend on σ, σ_W and σ_H. It was shown that the update rules for the optimization problem in (3.6) are given by [4],

$$\begin{cases} H_{ij} \leftarrow H_{ij} \frac{(W^T V)_{ij}}{(W^T WH + \beta H)_{ij}} \\ \\ W_{ij} \leftarrow W_{ij} \frac{(VH^T)_{ij}}{(WHH^T + \alpha W)_{ij}}. \end{cases} \qquad (3.7)$$

Observe that, since the data matrix V is non-negative, the update rules in (3.7) lead to non-negative factors W and H as long as the initial values of the algorithm are chosen to be non-negative.

3.2 NMF as a Clustering Technique

In [5], NMF was used to extract relevant biological correlations in gene expression data. The data was represented by an expression matrix A of size $N \times M$, whose rows contain the expression levels of N genes in M samples. The NMF reduces the dimensionality of the gene expression data into a small number ($k < N$) of "metagenes", defined as positive linear combinations of the N genes. Then, the gene expression pattern of the samples are approximated as positive linear combinations of these metagenes. Mathematically, this corresponds to factoring the matrix A into two matrices with positive entries, $A \approx WH$.

Matrix W has size $N \times k$, with each of the k columns defining a metagene; The w_{ij} entry is the coefficient of gene i in metagene j. Matrix H has size $k \times M$, with each of the M columns representing the metagene expression pattern of the corresponding sample; The h_{ij} entry represents the expression level of metagene i in sample j. Given a factorization $A \approx WH$, the matrix H is used to group M samples into k clusters. Each sample is placed into a cluster corresponding to the most highly expressed metagene in the sample; that is, sample j is placed in cluster i if the h_{ij} is the largest entry in column j; (see Fig. (1)). In [24], Lee *et al.* used the NMF to decompose images of human faces into
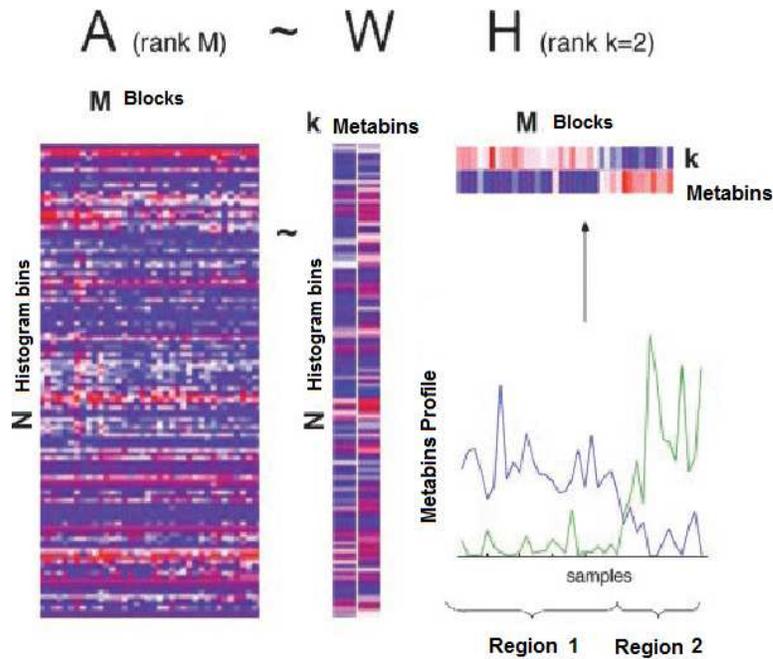
Figure 1. Rank-2 reduction of a DNA microarray of N genes and M samples is obtained by NMF, $A \approx WH$. Metagene expression levels (rows of H) are color coded by using a heat color map, from dark blue (minimum) to dark red (maximum). The same data is shown as continuous profiles below. The relative amplitudes of the two metagenes determine two classes of samples, class 1 and class 2. Here, the samples were ordered to better expose the class distinction [5].

parts reminiscent of features such as eyes and noses, while the application of traditional factorization methods, such as principal component analysis (PCA) and vector quantization (VQ), to the image data yielded components with no obvious interpretation. The database of images is viewed as an $n \times m$ matrix V, where each column contains n non-negative pixel values of one of the m facial images. The NMF leads to the decomposition $V \approx WH$, where the dimensions of the factors W and H are $n \times r$ and $r \times m$, respectively. The r columns of W are termed "basis images". Each column of H is called an *encoding* and is in one-to-one correspondence with a face in V. An encoding consists of the coefficients by which a face is represented as a linear combination of basis images. The NMF basis and encodings contain a large fraction of vanishing coefficients. Both the basis images and the image encodings are sparse. The basis images are sparse because they are "non-global" and contain several versions of mouths, noses and other facial parts, where the various versions are in different locations or forms. The variability of a whole face is generated by combining these different parts. Although all parts are used by at least one face, any given face does not use all the available parts. This results in a sparsely distributed image encoding, in contrast to the unary encoding of VQ and the fully distributed PCA encoding [24] as shown in Fig. (*2*).

In the sequel of this thesis, we will show how to leverage NMF for image segmentation.
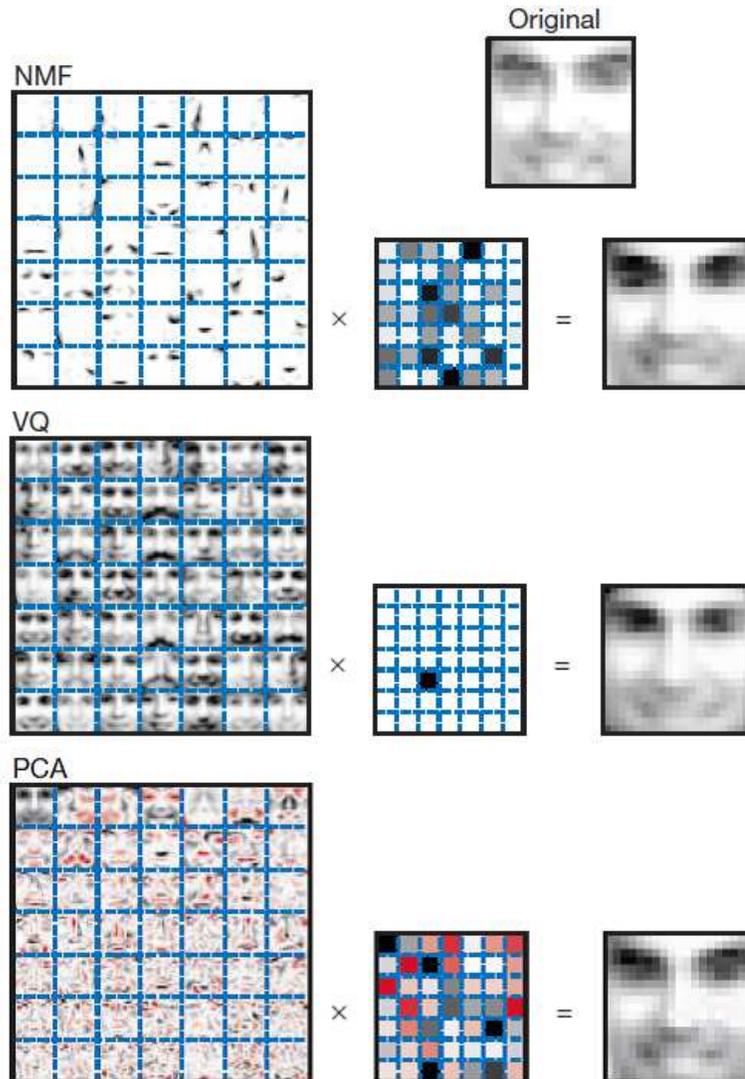
Figure 2. Non-negative matrix factorization (NMF) learns a parts-based representation of faces, whereas vector quantization (VQ) and principal component analysis (PCA) learn holistic representations. The three learning methods find approximate factorizations of the form $V \approx WH$, but with three different types of constraints on W and H. As shown in the 7×7 montages, each method has learned a set of $r = 49$ basis images. Positive values are illustrated with black pixels and negative values with red pixels. A particular instance of a face, shown at top right, is approximately represented by a linear superposition of basis images. The coefficients of the linear superposition are shown next to each montage, in a 7×7 grid, and the resulting superpositions are shown on the other side of the equality sign. Unlike VQ and PCA, NMF learns to represent faces with a set of basis images resembling parts of faces, such as eyes, mouth and nose [24].

Chapter 4

Deformable Models

In this chapter, we explain the detailed theoretical and mathematical formulation of the two classes of deformable models: the parametric deformable model, also known as *snake* or *active contour*, and the geometric deformable model, also known as the *level set method*.

Deformable models refer to a powerful class of physics-based modeling techniques widely employed in the image synthesis, image analysis, image segmentation, shape design and related fields. By numerically simulating the governing equations of motion, typically expressed as Partial Differential Equations (PDEs) in a continuous setting, deformable models mimic various generic behaviors of natural non-rigid materials in response to applied forces, such as continuity, smoothness and elasticity. Deformable models offer a potent approach that combines geometry, physics and approximation theory. Such models can be used to infer image disparity fields, image flow fields, and to infer the shapes and motions of objects from static or video images. In this context, deformable models are subject to external forces that impose constrains derived from image data. The forces actively shape and move the model to achieve maximal consistency with the objects of interest and maintain consistency over time. These models are widely used in medical image analysis; they have proven to be very useful in segmenting, matching and tracking anatomic structures by exploring constraints derived from the image data in conjunction with a priori knowledge about the location, size and shape of these structures [32]. There are two types of deformable models: parametric and geometric models. Parametric deformable models represent curves and surfaces explicitly in its parametric form, i.e., using a set of contour points. Its popularity in medical image analysis is credited to the work of snake or (active

contour) by [20]. The geometric deformable model (level set method) is based on the theory of curve evolution and geometric flow, which represents curves and surfaces implicitly as a level set of an evolving higher dimensional function [19].

4.1 The Parametric Deformable Model (Snake or Active Contour)

Active contour models (ACMs) are based on the idea of evolving a curve in the image domain under the influence of an internal energy and an external energy [20]. The optimal contour minimizes the total energy functional, given by the weighted sum of the internal energy and external energy terms. This curve is represented in a parameterized form by a set of contour points. The internal energy defines the shape of the contour and imposes smoothness and relevant geometrical constraints on the curve, e.g., Eq. (4.1). In Equation (4.1), $v(s)$ denotes the parameterized curve, where the points $(x(s), y(s))$ move through the spatial domain of the image to minimize the energy functional, and w_1, w_2 are the weighting parameters that control the contour's elasticity and rigidity. The external energy in Eq. (4.2) is computed from the image and attracts the contour towards objects boundaries and other desired salient features in the image. The total active contour energy functional is given in Eq. (4.3). This model is used for image segmentation and some other applications in image processing, such as edge detection, shape modeling and motion tracking [20].

$$E^{Internal} = \frac{w_1(s)}{2}|v'(s)|^2 + \frac{w_2(s)}{2}|v''(s)|^2, \tag{4.1}$$

$$E^{External} = -|\nabla I(v(s))|^2, \tag{4.2}$$

$$E^{ACM} = E^{Internal} + E^{External} = \frac{w_1(s)}{2}|v'(s)|^2 + \frac{w_2(s)}{2}|v''(s)|^2 - |\nabla I(v(s))|^2. \tag{4.3}$$

In order to minimize the total energy E^{ACM}, we formulate this functional as a Euler-Lagrange functional $L(s, \upsilon(s), \upsilon'(s), \upsilon''(s))$ and solve the Euler-Lagrange equation of the general form:

$$F(f) = \int_{x_0}^{x_1} L(x, f(x), f'(x), f''(x), ..., f^n(x))dx, \ f' = \frac{df}{dx}, f'' = \frac{d^2 f}{dx^2}, f^n = \frac{d^n f}{dx^n}.$$

(4.4)

The stationary values of the functional $F(f)$ in Eq. (4.4) can be obtained by solving the Euler-Lagrange equation [8]:

$$\frac{\partial L}{\partial f} - \frac{d}{dx}(\frac{\partial L}{\partial f'}) + \frac{d^2}{dx^2}(\frac{\partial L}{\partial f''}) - .... + (-1)^n \frac{d^n}{dx^n}(\frac{\partial L}{\partial f^n}) = 0.$$

(4.5)

By adding the time variable, it can be shown that minimizing E^{ACM} corresponds to solving the following Euler-Lagrange equation.

$$\upsilon_t(s, t) = -(w_1 \upsilon')' + (w_2 \upsilon'')'' + \nabla E^{External}(\upsilon) = 0,$$

(4.6)

where $\upsilon_t = \frac{\partial \upsilon}{\partial t}$. The finite difference equation in (4.6) can be solved numerically using gradient descent.

4.2 Drawbacks of the Active Contour Model

There are two main difficulties in the active contour model: First, the contour must be initialized fairly close to the final target in order to converge. However, to make a "good" initialization, we need to have a "good" estimate of the solution before starting the iterative process of adapting the contour. This leads to a solution that is sensitive to the initial condition. Secondly, there are difficulties associated with the topological changes for the merging and splitting of the evolving curve. These difficulties lie in the parametric representation of the contour. For instance, when the contour merges and splits to fit the objects'

boundaries in the image, one has to keep track of which points are in which contour and what their order is. Figure *3* shows an example, where the active contour model fails in handling the topological changes associated with the movement of the fingers [13].

Figure 3. The active contour model fails to handle the topological changes of the moving fingers [13].

4.3   The Level Set Method

The basic idea of the level set method [33], [31] is to embed the moving contour as the zero iso-contour of a higher-dimension implicit function $\phi : R^n \times R^+ \longrightarrow R$. In the 2D spatial dimension, the closed curve, denoted by C, can be implicitly represented as the zero level $C(t) = \{(x, y)|\phi(x, y, t) = 0\}$ of a level set function $\phi(x, y, t)$. The evolution of the level

set function follows the Hamilton- Jacobi equation:

$$\begin{cases} \frac{\partial \phi}{\partial t} + F|\nabla \phi| = 0 \\ \\ \phi(x, y, 0) = \phi_0(x, y), \end{cases} \tag{4.7}$$

where the function F, called the *speed function*, is defined depending on the mean-curvature [33] or image edges information [6]. In the early implementations of the level set method [30], [31], the level set function ϕ can develop discontinuous jumps, known as *shocks*, during the evolution, which makes further computation highly inaccurate. To avoid these problems, a common numerical scheme is to initialize the function ϕ as a signed distance function before the evolution, and then "reshape" (or "re-initialize") the function ϕ to be a signed distance function periodically during the evolution. It is crucial to keep the evolving level set function as an approximate signed distance function during the evolution, especially in a neighborhood around the zero level set. It is well known that a signed distance function must satisfy a desirable property of $|\nabla \phi| = 1$. Conversely, any function ϕ satisfying $|\nabla \phi| = 1$ is the signed distance function plus a constant. In order to avoid the drawbacks of the re-initialization procedure, Li *et al* proposed in [27] a variational level set formulation that does not require o the re-initialization procedure by using the following integral:

$$R(\phi) = \int_\Omega 0.5(|\nabla \phi| - 1)^2 dx, \tag{4.8}$$

as a metric to characterize how close a function ϕ is to a signed distance function. Then the level set variational formulation is given by:

$$E(\phi) = \mu R(\phi) + E_m(\phi), \tag{4.9}$$

where $\mu > 0$ is a parameter controlling the effect of penalizing the deviation of ϕ from a signed distance function, and $E_m(\phi)$ is a certain energy that would drive the motion of the zero level curve of ϕ.

Using calculus of variation, the evolution of the level set function is given by:

$$\frac{\partial \phi}{\partial t} = -\frac{\partial E}{\partial \phi}. \qquad (4.10)$$

For the purpose of image segmentation, E_m is defined as a functional that depends only on the image data, and is, therefore, called the *external energy*. Accordingly, the energy $R(\phi)$ is called the *internal energy* of the function ϕ, since it is a function of ϕ only [27]. We can also add other smoothing constraints to the energy functional E.

Chapter 5

NMF-Based Level Set Segmentation

In this chapter, we will show how the NMF can be used to discover the image regions and estimate their number. We will present the theoretical and visual interpretations of the W, and H factors of the NMF and explain how they can be used to build a robust energy functional of the image. The proposed NMF-based level set model is subsequently derived.

5.1 Region Discovery using NMF

We first construct a data matrix based on the histogram of the image, rather than the intensity values directly. This data matrix takes into account the local information in the image by partitioning the image into m blocks and computing the histogram of each block. The histograms of the blocks are then stacked to form the columns of the data matrix V. The data matrix $V = \{v_{ij}\}$ is an $n \times m$ matrix, where n is the number of intensity bins or ranges in the histograms standardized for all image blocks. Specifically, the (i, j) entry, v_{ij}, is the number of pixels in the block j with intensity range in the bin i. The rows of V describe the ranges of intensity in the bin i in every j block. Our goal is to find $k < m$ "basic histograms" such that the histogram of every image block can be expressed as a positive linear combination of the basic histograms. This can be achieved using non-negative matrix factorization (NMF). NMF provides a natural way to cluster the histogram data matrix, because it involves nonnegative entries. Other matrix decomposition techniques, such as principal component analysis (PCA) or singular value decomposition (SVD) do not guarantee the non-negativity constraints, and hence loose the physical and intuitive interpretation of the factorization. However, this non-negativity requirement makes the factorization

problem more challenging, as we saw in Chapter 3. We use the Probabilistic NMF (PNMF) algorithm in [4], which takes into account the noise in the data matrix and performs a maximum a posteriori NMF factorization. This algorithm was presented in Chapter 3 Section 3.1.3.

Mathematically, the task of finding the basic histograms of the image corresponds to factoring the histogram data matrix V into two matrices with positive entries $V \approx WH$, where W is $n \times k$ and H is $k \times m$. The k columns of W define the basic histograms, which we will show, correspond to the histograms of the distinct image regions. The k rows of the matrix H cluster the data matrix into metabins, and represents the distribution of the image regions within the m blocks. An illustration of the NMF factorization of the data matrix is provided in Fig. (*4*).

The PNMF factorization $V \approx WH$ induces then a clustering of the histogram data matrix into k basic histograms or k metabin regions $\Omega_i, i = 1, \cdots, k$. In the sequel, we will investigate how the non-negative matrices W and H provides statistical and spatial information about the clustered regions in the image. We first consider the synthetic binary image in Fig. (*5*). The PNMF of the data matrix of this image with $k = 2$ results in the W and H matrices shown in Fig. (*5a*), (*5b*), respectively. Plotting the entries of each column of W, we obtain two sharp peaks: one peak at the $(0 - 1)$ range of intensity value, corresponding to the black region, and a second peak at the $(254 - 255)$ range, corresponding to the white region. Hence, the matrix W seems to provide the distribution of the pixel intensity values in each region, and from this distribution, we can obtain the statistical information (mean and variance of every region in the image). The normalized entries of the columns of H provide the percentage of pixels in every block that are within

27
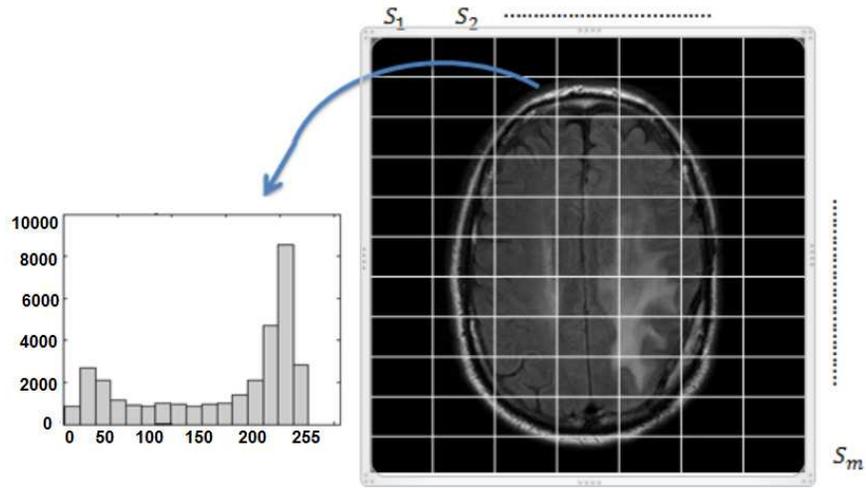
(a) Dividing the image into blocks and computing the histogram of each block.
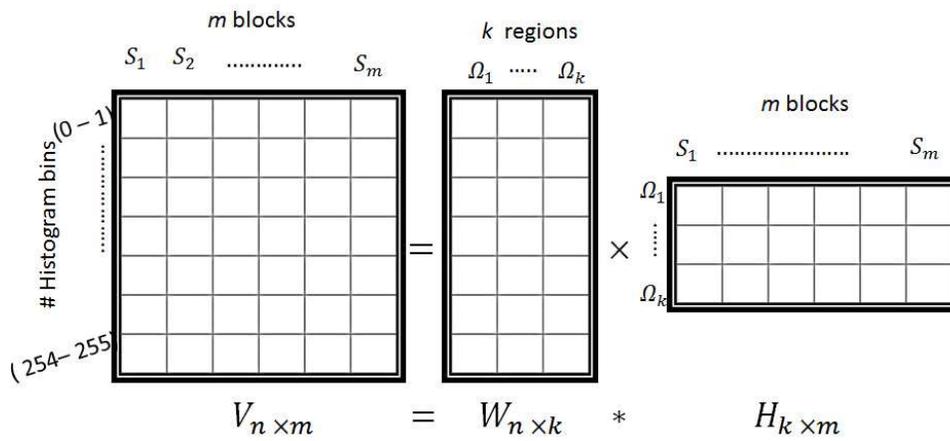
(b) Data matrix and histogram factorization.

Figure 4. Building the data matrix and histogram factorization using PNMF.

the clustered regions of the image. For instance, when the image block is included entirely in one region, we obtain the value of 1 in the entry that corresponds to that region and zero in the other. While, if two regions are included in the block, we obtain the exact percentages of the local areas of these regions in that block (see Fig. (5b)).

The same interpretation has been reached on the synthetic gray-scale image given in Fig. (6). For this synthetic image, we started by meshing the image into $N \times N$ blocks first ($N = 16$ in this case). From the plot of the columns of the matrix W in Fig. (7a), we observe that the PNMF captured the four large regions in the image, while the two small regions in the image with intensity values (102) and (26) were not detected. Both W and H matrices were unable to capture these two small regions with a block size of (16×16). Even when we increase the number of regions k, we get additional zero columns in W, and additional zero rows in H, thus unable to capture these two small regions. However, by decreasing the size of the blocks to (8×8), so that the block size fits into these small regions, we found that the W and H matrices of the PNMF factorization specified six regions including the two small regions, as shown in Fig. (8). Hence, by decreasing the block size to partition the image and build the data matrix V, the PNMF resolution ability to distinguish small regions increases. The resolution of the PNMF is thus directly related to the block size used when partitioning the image. In other words, the smallest distinct region that can be detected by the PNMF has a size approximately equal to the block size.

In summary, applying the probabilistic non-negative matrix factorization PNMF on the histogram data matrix, that we build in section 5.1, provides two positive matrices that provide the statistical and spatial characteristics of the image regions. The matrix W provides the histogram distribution of each region in the image, which means that we can obtain the
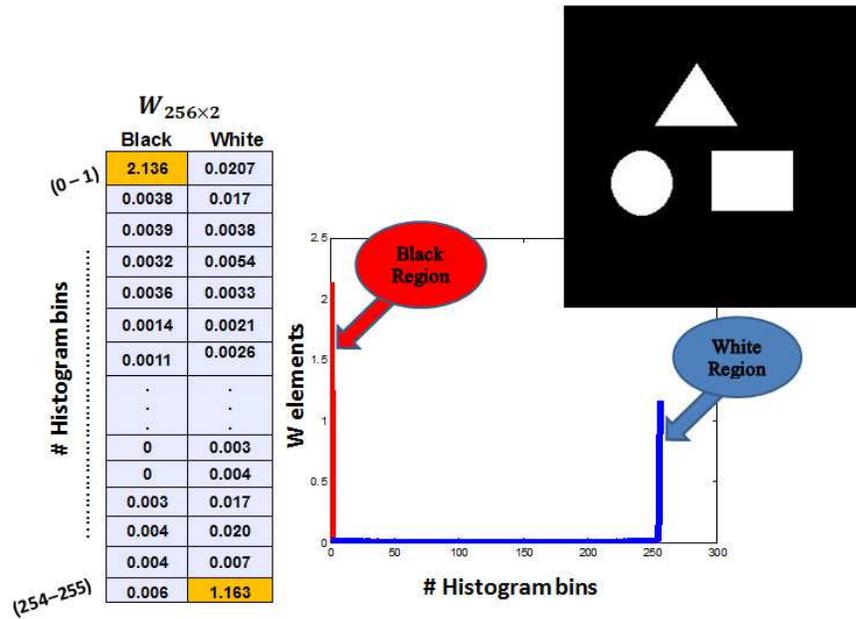
(a) "W" matrix Interpretation.
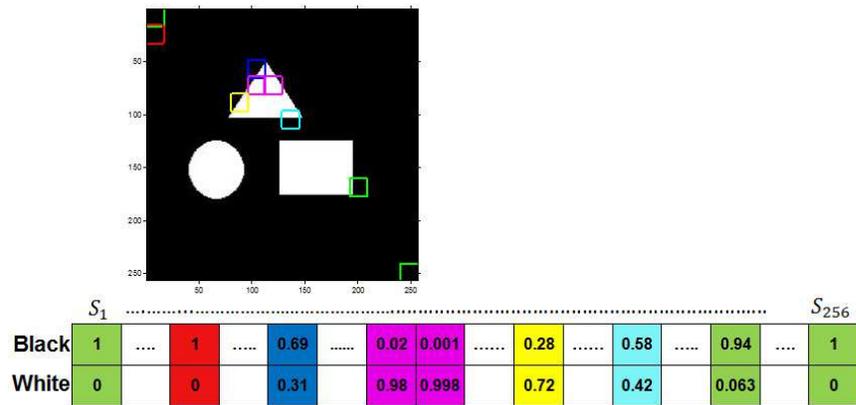
(b) "H" matrix interpretation.

Figure 5. "W" and "H" matrices interpretation for a synthetic binary image with 16×16 blocks and the entries of H normalized for each column.

statistical mean and variance of each region; While the matrix H provides the local spatial characteristics of each region: the percentage area of each region that is included in every block size. Based on these statistical and spatial characteristics of the image regions, we propose to build a robust external energy or a data term that will be used in the level set approach.

5.1.1 Estimating the number of distinct regions in the image. We show how to use the factor matrices W and H to estimate the number of distinct clusters or regions k. To illustrate the idea, we start by changing the value of k for the synthetic gray-scale image in Fig. (*6*) and observing the corresponding changes in the matrices W and H. We notice that when we increase k to be more than the true number of regions in the image (which is in this case $k = 6$), we obtain additional zero rows in H and additional zero columns in W. We, therefore, propose to use the sum of the *nuclear norms* (also known as the trace norms) of W and H. The nuclear norm of the matrix A is defined as

$$\|A\|_* = \text{trace}(\sqrt{A^H A}) = \sum_{i=1}^{\min\{m,n\}} \sigma_i, \tag{5.1}$$

where the σ_is are the singular values of the matrix A. We start by choosing an initial guess for the number of regions k_0 and we apply the PNMF on the histogram data matrix V with $k = k_0$. We obtain $V_{n \times m} \approx W_{n \times k_0} H_{k_0 \times m}$. We compute the sum of the nuclear norms of $W_{n \times k_0}$ and $H_{k_0 \times m}$. Then, we increase k and repeat the same steps. The sum nuclear norm is a nondecreasing function of k. This function platforms when $k \geq k^*$. The optimal number of regions is then given by k^*.

Fig. (*9*) shows how the sum of the nuclear norms increases with k for the synthetic gray-scale image in Fig. (*6*), until it stabilizes at $k = 6$, which corresponds to the exact

number of regions in the image.

The idea behind using the nuclear norm is that appending the matrix W, having singular values $\{\sigma_i\}_{i=1}^k$, by a column of zeros, $\mathbf{z}_w$, and forming the matrix $\tilde{W}_a = [W, \mathbf{z}_w]$ will add a zero singular value. In other words, the singular values of $\tilde{W}_a$ are given by $\{\{\sigma_i\}_{i=1}^k, 0\}$. Thus, the nuclear norm does not change when we add a column of zeros. Similarly, appending the matrix H by a row of zeros, $\mathbf{z}_h^t$, and forming the matrix $\tilde{H}_a = \begin{bmatrix} H \\ \mathbf{z}_h^t \end{bmatrix}$, will not change the sum of the singular values. In practice, we do not have exact zeros in the additional columns and rows of the appended matrices $\tilde{W}_a$ and $\tilde{H}_a$, respectively, but we have small values that are close to zero. We will show that if the column vector $\mathbf{z}_w$ and the row vector $\mathbf{z}_h^t$ have small entries, then the additional singular values in the appended matrices, $\tilde{W}_a$ and $\tilde{H}_a$, are also small. In particular, the nuclear norm will change only slightly when we append a column or a row of small values. The following lemma formalizes and proves this result.

Lemma 1. *Let $\boldsymbol{W} \in \mathbb{R}^{n \times k}$ and $\boldsymbol{H} \in \mathbb{R}^{k \times m}$, where $k < \min\{n, m\}$. Let $\{\sigma_i^W\}_{i=1}^k$ and $\{\sigma_i^H\}_{i=1}^k$ be the singular values of $\boldsymbol{W}$ and $\boldsymbol{H}$, respectively. Consider the augmented matrices $\boldsymbol{W}_a = [\boldsymbol{W}|\boldsymbol{0}]$ and $\boldsymbol{H}_a = \left[\begin{smallmatrix} \boldsymbol{H} \\ \boldsymbol{0}^t \end{smallmatrix}\right]$, where $\boldsymbol{0}$ denotes the vector with zero entries. Then*

$$\|\boldsymbol{W}_a\|_* = \|\boldsymbol{W}\|_* \quad and \quad \|\boldsymbol{H}_a\|_* = \|\boldsymbol{H}\|_*. \tag{5.2}$$

Similarly, consider the augmented matrices $\tilde{\boldsymbol{W}}_a = [\boldsymbol{W}|\boldsymbol{z}_w]$ and $\tilde{\boldsymbol{H}}_a = \left[\begin{smallmatrix} \boldsymbol{H} \\ \boldsymbol{z}_h^t \end{smallmatrix}\right]$, where $\boldsymbol{z}_w$ and $\boldsymbol{z}_h$ are vectors. Then, we have

$$\|\boldsymbol{W}\|_* \leq \|\tilde{\boldsymbol{W}}_a\|_* \leq \|\boldsymbol{W}\|_* + \|\boldsymbol{z}_w\|, \tag{5.3}$$

$$\|\boldsymbol{H}\|_* \leq \|\tilde{\boldsymbol{H}}_a\|_* \leq \|\boldsymbol{H}\|_* + \|\boldsymbol{z}_h\|. \tag{5.4}$$

In particular, if $\|\boldsymbol{z}_w\| \leq \varepsilon$, and $\|\boldsymbol{z}_h\| \leq \varepsilon'$, then

$$0 \leq \|\tilde{\boldsymbol{W}}_a\|_* - \|\boldsymbol{W}\|_* \leq \varepsilon, \tag{5.5}$$

$$0 \leq \|\tilde{\boldsymbol{H}}_a\|_* - \|\boldsymbol{H}\|_* \leq \varepsilon'. \tag{5.6}$$

5.1.2 Proof of Lemma 1. We will first prove that appending the matrix $\boldsymbol{W}$ by a column of zeros and the matrix $\boldsymbol{H}$ by a row of zeros will not change the sum of the singular values of the appended matrices $\boldsymbol{W}_a$ and $\boldsymbol{H}_a$.

We have,

$$\{\sigma_i^{\boldsymbol{W}_a}\}_{i=1}^k = \{\lambda_i(\boldsymbol{W}_a^t \boldsymbol{W}_a)\}_{i=1}^k, \tag{5.7}$$

where $\{\lambda_i(\boldsymbol{W}_a^t \boldsymbol{W}_a)\}_{i=1}^k$ are the eigenvalues of the matrix $\boldsymbol{W}_a^t \boldsymbol{W}_a$,

$$\boldsymbol{W}_a^t \boldsymbol{W}_a = \left[\begin{smallmatrix} \boldsymbol{W}^t \\ \boldsymbol{0}^t \end{smallmatrix}\right] * [\boldsymbol{W}|\boldsymbol{0}] = \left[\begin{array}{c:c} \boldsymbol{W}^t \boldsymbol{W} & \boldsymbol{0} \\ \hdashline \boldsymbol{0}^t & 0 \end{array}\right]; \tag{5.8}$$

To find the eigenvalues $\{\lambda_i(\boldsymbol{W}_a^t \boldsymbol{W}_a)\}_{i=1}^k$ we need to solve the characteristic polyno-

mial:

$$\det(\boldsymbol{W}_a^t \boldsymbol{W}_a - \lambda \boldsymbol{I}) = -\lambda \det(\boldsymbol{W}^t \boldsymbol{W} - \lambda \boldsymbol{I}) = 0. \tag{5.9}$$

Then

$$\lambda(\boldsymbol{W}_a^t \boldsymbol{W}_a) = \{\lambda(\boldsymbol{W}^t \boldsymbol{W}), 0\} \Rightarrow \sigma(\boldsymbol{W}_a) = \{\sigma(\boldsymbol{W}), 0\}. \tag{5.10}$$

Similarly, we can show that if $\boldsymbol{H}_a = \left[\begin{smallmatrix} H \\ \mathbf{0} \end{smallmatrix}\right]$, then $\lambda(\boldsymbol{H}_a \boldsymbol{H}_a^t) = \{\lambda(\boldsymbol{H} \boldsymbol{H}^t), 0\} \Rightarrow \sigma(\boldsymbol{H}_a) = \{\sigma(\boldsymbol{H}), 0\}$.

We now prove the second part of the Lemma, namely, Eqs. (5.3) and (5.4). First, we can write the appended matrix $\tilde{\boldsymbol{W}}_a$ as follows:

$$\tilde{\boldsymbol{W}}_a = [\boldsymbol{W} | \mathbf{z}_w] = [\boldsymbol{W} | \mathbf{0}_{n \times 1}] + [\mathbf{0}_{n \times k} | \mathbf{z}_w], \tag{5.11}$$

where the first zero in $[\boldsymbol{W} | \mathbf{0}_{n \times 1}]$ is a $n \times 1$ vector, while the second zero in $[\mathbf{0}_{n \times k} | \mathbf{z}_w]$ is a matrix with the same dimension as $\boldsymbol{W}$. Using the triangular inequality, we have

$$\|\boldsymbol{W} | \mathbf{z}_w\|_* \leq \|\boldsymbol{W} | \mathbf{0}\|_* + \|\mathbf{0} | \mathbf{z}_w\|_*. \tag{5.12}$$

We can easily see that $\|\mathbf{0}_{n \times k} | \mathbf{z}_w\|_* = \|\mathbf{z}_w\|$ and from the first part of the proof, $\|\boldsymbol{W} | \mathbf{0}_{n \times 1}\|_* = \|\boldsymbol{W}\|_*$. Thus,

$$\|\boldsymbol{W} | \mathbf{z}_w\|_* \leq \|\boldsymbol{W}\|_* + \|\mathbf{z}_w\|. \tag{5.13}$$

In particular, small values of $\mathbf{z}_w$ result in a small perturbation of the nuclear norm of $\tilde{\boldsymbol{W}}_a$.

In order to prove the left hand side of the Eqs. (5.3), (5.4) we use the determinant formula for block matrices $\det \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} = \det(\boldsymbol{A}) \det(\boldsymbol{D} - \boldsymbol{C} \boldsymbol{A}^{-1} \boldsymbol{B})$. We have

$$\tilde{\boldsymbol{W}}_a^t \tilde{\boldsymbol{W}}_a - \lambda \boldsymbol{I} = \left[\begin{smallmatrix} \boldsymbol{W}^t \\ \mathbf{z}_w^t \end{smallmatrix}\right] * [\boldsymbol{W} | \mathbf{z}_w] - \lambda \boldsymbol{I} = \left[\begin{array}{c:c} \boldsymbol{W}^t \boldsymbol{W} - \lambda \boldsymbol{I} & \boldsymbol{W}^t \mathbf{z}_w \\ \hdashline \mathbf{z}_w^t \boldsymbol{W} & \|\mathbf{z}_w\|^2 - \lambda \end{array}\right]. \tag{5.14}$$

34

Using the determinant formula we can write the determinant of $\tilde{\boldsymbol{W}}_a^t \tilde{\boldsymbol{W}}_a - \lambda \boldsymbol{I}$ as follows

$$\det \left(\boldsymbol{W}^t \boldsymbol{W} - \lambda \boldsymbol{I}\right) \det \left(\|\mathbf{z}_w\|^2 - \lambda - \mathbf{z}_w^t \boldsymbol{W}(\boldsymbol{W}^t \boldsymbol{W} - \lambda \boldsymbol{I})^{-1} \boldsymbol{W}^t \mathbf{z}_w\right) = 0, \qquad (5.15)$$

$$\Rightarrow \left\{ \begin{array}{l} \det \left(\boldsymbol{W}^t \boldsymbol{W} - \lambda \boldsymbol{I}\right) = 0, \\[2mm] \det \left(\|\mathbf{z}_w\|^2 - \lambda - \mathbf{z}_w^t \boldsymbol{W}(\boldsymbol{W}^t \boldsymbol{W} - \lambda \boldsymbol{I})^{-1} \boldsymbol{W}^t \mathbf{z}_w\right) = 0 \Rightarrow \lambda^* \geq 0. \end{array} \right\}. \qquad (5.16)$$

Hence, the eigenvalues of $\tilde{\boldsymbol{W}}_a^t \tilde{\boldsymbol{W}}_a$ are $\{\lambda(\boldsymbol{W}^t \boldsymbol{W}), \lambda^*\}$, where $\lambda(\boldsymbol{W}^t \boldsymbol{W})$ are the eigenvalues of $\boldsymbol{W}^t \boldsymbol{W}$ and $\lambda^* \geq 0$ because $\boldsymbol{W}^t \boldsymbol{W}$ is a positive semi-definite matrix.. Then

$$\|\tilde{\boldsymbol{W}}_a\|_* = \sum_i \sigma_i(\tilde{\boldsymbol{W}}_a) = \sum_i \sigma_i(\boldsymbol{W}) + \sqrt{\lambda^*} \Rightarrow \|\tilde{\boldsymbol{W}}_a\|_* \geq \|\boldsymbol{W}\|_*. \qquad (5.17)$$

Similarly, for the matrix $\boldsymbol{H}$, we can use the triangular inequality and the determinant formula of the block matrices to prove the following:

$$\|\boldsymbol{H}\|_* \leq \|\tilde{\boldsymbol{H}}_a\|_* \leq \|\boldsymbol{H}\|_* + \|\boldsymbol{z}_h\|. \qquad (5.18)$$

5.2   Proposed Variational Framework

We consider the image model $I(\boldsymbol{x}) = J(\boldsymbol{x}) * b(\boldsymbol{x}) + n(\boldsymbol{x})$, where $I(\boldsymbol{x})$ is the observed intensity at pixel $\boldsymbol{x}$, $J(\boldsymbol{x})$ is the "true" (noiseless/unbiased) intensity at pixel $\boldsymbol{x}$, $b(\boldsymbol{x})$ is the bias field associated with $\boldsymbol{x}$ and $n(\boldsymbol{x})$ is the observation noise. We take into account the two matrices W and H in the PNMF factorization $V \approx WH$ to build the proposed energy functional. This functional codes the external energy and contains two terms: a statistical term and a spatial term. The statistical term uses the matrix W and characterizes the mean and standard deviation of the histogram of each region. The spatial term relies on the matrix H and characterizes the local spatial area of each region inside the blocks. In order

to compute the statistical energy term, we formulate the segmentation problem as one of computing the maximum a posteriori (MAP) partition of the image domain Ω into disjoint regions by maximizing the posterior probability $p(\{\Omega_1, \Omega_2, \cdots, \Omega_k\}|I)$ for the image I. According to the Bayes rule

$$p(\{\Omega_1, \Omega_2, \cdots, \Omega_k\}|I) \propto p(I|\{\Omega_1, \Omega_2, \cdots, \Omega_k\}) \, p(\{\Omega_1, \Omega_2, \cdots, \Omega_k\}). \qquad (5.19)$$

Assuming that the prior probabilities of all partitions $p(\{\Omega\})$ are equal, and the pixels within each region are independent, the MAP estimate reduces to finding the maximum of $\prod_{i=1}^{k} \prod_{\boldsymbol{x} \in \Omega_i} p_i(I(\boldsymbol{x}))$, where $p_i(I(\boldsymbol{x})) = p(I(\boldsymbol{x})|\Omega_i)$, $i = 1, 2, ..., k$. By taking the logarithm, the maximization can be converted to the minimization of the following energy function:

$$E^{Statistical} = \sum_{i=1}^{k} \int_{\Omega_i} - \log p_i(I(\boldsymbol{x}))d\boldsymbol{x}, \qquad (5.20)$$

where $p_i(I(\boldsymbol{x}))$ is modeled as a Gaussian distribution.

$$p_i(I(\boldsymbol{x})) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp(-\frac{(I(\boldsymbol{x}) - \mu_i b(\boldsymbol{x}))^2}{2(\sigma_i)^2}), \qquad (5.21)$$

where μ_i, and σ_i are computed form the matrix W. To compute the spatial energy term, we consider the matrix H, which induces a local spatial clustering of the regions Ω_i, $i = 1, \cdots, k$ in each block. Ideally, if the block j is included entirely in the region i, then $h_{ij} = 1$, and $h_{lj} = 0$ for $l \neq i$. Hence, by dividing the entries of H by the sum of each column, we can interpret h_{ij} as the proportion of the area of region i in the block j. Therefore, we can represent the local spatial area of each region i inside the block j as a weighted linear combination of the block area where the weights are given by the normalized entries of the jth column of the matrix H. We propose the following spatial

data term:

$$E^{Spatial} = \sum_{i=1}^{k} \sum_{j=1}^{m} \left(\int_{\Omega_i} \mathcal{I}_{S_j}(\boldsymbol{x}) d\boldsymbol{x} - \frac{h_{ij}a}{\sum_{i=1}^{k} h_{ij}} \right)^2, \tag{5.22}$$

where a is the area of each block (all blocks are assumed to have equal areas), and $\mathcal{I}_{S_j}(\boldsymbol{x})$ is the characteristic function that introduces all pixels in the block S_j, and is defined as follows:

$$\mathcal{I}_{S_j}(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \boldsymbol{x} \in S_j \\ 0, & \text{otherwise.} \end{cases} \tag{5.23}$$

The total data term is then given by the sum of the statistical energy and the spatial energy terms.

$$E = E^{Statistical} + E^{Spatial}, \tag{5.24}$$

$$E = \sum_{i=1}^{k} \left[\int_{\Omega_i} (\log(\sqrt{2\pi}\sigma_i) + \frac{(I(\boldsymbol{x}) - \mu_i b(\boldsymbol{x}))^2}{2\sigma_i^2}) d\boldsymbol{x} + \sum_{j=1}^{m} \left(\int_{\Omega_i} \mathcal{I}_{S_j}(\boldsymbol{x}) d\boldsymbol{x} - \frac{h_{ij}a}{\sum_{i=1}^{k} h_{ij}} \right)^2 \right]. \tag{5.25}$$

The energy functional E is subsequently converted to a level set formulation by generating the level set functions $\phi(\boldsymbol{x})$ and representing the disjoint regions with a number of membership functions $M_i(\phi(\boldsymbol{x}))$. The membership functions satisfy two constraints: i) they are valued in $[0, 1]$ and ii) the summation of all membership functions is equal to 1, i.e., $\sum_{i=1}^{k} M_i(\phi(\boldsymbol{x})) = 1$. This can be achieved by representing the membership function as a smoothed version of the Heaviside function. For example, in the two-phase formulation, the regions Ω_1 and Ω_2 can be represented with their membership functions defined by $M_1(\phi) = H(\phi)$ and $M_2(\phi) = 1 - H(\phi)$ respectively, where H is the Heaviside function. For a multi-phase formulation, the combination of the Heaviside functions is different. For example, in the four-phase formulation, we have two level set functions ϕ_1 and ϕ_2. The

membership functions are given as follows: $M_1 = H(\phi_1)H(\phi_2)$, $M_2 = H(\phi_1)(1-H(\phi_2))$, $M_3 = (1 - H(\phi_1))H(\phi_2)$ and $M_4 = (1 - H(\phi_1))(1 - H(\phi_2))$. The total energy in Eq. (5.25) can be equivalently expressed as the following level set energy functional:

$$E(\phi, b) = \sum_{i=1}^{k} \left[\int_{\Omega} \left(\log(\sqrt{2\pi}\sigma_i) + \frac{(I(\boldsymbol{x}) - \mu_i b(\boldsymbol{x}))^2}{2\sigma_i^2} \right) M_i(\phi(\boldsymbol{x}))d\boldsymbol{x} \right]$$
$$+ \sum_{i=1}^{k} \sum_{j=1}^{m} \left(\int_{\Omega} \mathcal{I}_{S_j}(\boldsymbol{x}) M_i(\phi(\boldsymbol{x}))d\boldsymbol{x} - \frac{h_{ij}a}{\sum_{i=1}^{k} h_{ij}} \right)^2. \tag{5.26}$$

Equation (5.26) can be rewritten as:

$$E(\phi, b) = \sum_{i=1}^{k} \left[\int_{\Omega} e_i(\boldsymbol{x}, b) M_i(\phi(\boldsymbol{x}))d\boldsymbol{x} + \sum_{j=1}^{m} \left(\int_{\Omega} \mathcal{I}_{S_j}(\boldsymbol{x}) M_i(\phi)d\boldsymbol{x} - \frac{h_{ij}a}{\sum_{i=1}^{k} h_{ij}} \right)^2 \right],$$
$$\tag{5.27}$$

where $e_i(\boldsymbol{x}, b) = \log(\sqrt{2\pi}\sigma_i) + \frac{(I(x) - \mu_i b(x))^2}{2\sigma_i^2}$. The energy term $E(\phi, b)$ represents the external energy or the data term in the total energy of the proposed variational level set formulation. The total external and internal energy is given by

$$\mathcal{F}(\phi, b) = \alpha E(\phi, b) + \beta R(\phi) + \gamma L_g(\phi) + \nu A_g(\phi), \tag{5.28}$$

where $R(\phi)$, $L_g(\phi)$ and $A_g(\phi)$ are the regularization terms, and α, β, γ and ν are weighting parameters. The energy term $R(\phi)$, defined by $R(\phi) = \frac{1}{2} \int_{\Omega} (|\nabla\phi| - 1)^2 d\boldsymbol{x}$, is a distance regularization term [28] that is minimized when $|\nabla\phi| = 1$, a property of the signed distance function. The second energy term, $L_g(\phi) = \int_{\Omega} g|\nabla H(\phi(\boldsymbol{x})|d\boldsymbol{x}$, computes the arc length of the zero level set contour, $(\int_{\Omega} |\nabla H(\phi(\boldsymbol{x})|d\boldsymbol{x})$, and therefore serves to smooth the contour by penalizing its arc length during propagation. The contour length is weighted by the edge indication function $g = \frac{1}{1+|\nabla(G_\sigma * I)|^2}$, where $G_\sigma * I$ is the convolution of the image I with the smoothing Gaussian kernel G_σ. The function g works to stop the level set evolution near the optimal solution; since it is near zero in the variational edges and positive otherwise.

Therefore, the regularization term L_g serves to minimize the length of the level set curve at the image edges. The third regularization term, $A_g(\phi) = \int_\Omega gH(\phi(\boldsymbol{x})d\boldsymbol{x}$, is the area obtained by the level set curve weighted by the edge indication function.

Finally, the total energy functional to be minimized for the purpose of segmentation is expressed as:

$$\mathcal{F}(\phi, b) = \alpha \sum_{i=1}^{k} \left[\int_\Omega e_i(\boldsymbol{x}, b) M_i(\phi) dx + \sum_{j=1}^{m} \left(\int_\Omega \mathcal{I}_{S_j}(\boldsymbol{x}) M_i(\phi) d\boldsymbol{x} - \frac{h_{ij} a}{\sum_{i=1}^{k} h_{ij}} \right)^2 \right]$$

$$+ \frac{\beta}{2} \int_\Omega (|\nabla\phi| - 1)^2 dx + \gamma \int_\Omega g|\nabla H(\phi|dx + \nu \int_\Omega gH(\phi)d\boldsymbol{x}, \tag{5.29}$$

5.3   Level Set Formulation and Energy Minimization

The minimization of the energy functional $\mathcal{F}$ in Eq. (5.29) can be achieved iteratively by minimizing $\mathcal{F}$ w.r.t. each of the two variables, ϕ_l and b, assuming that the other variable is constant. We first fix the variable b, then the minimization of the energy functional $\mathcal{F}(\phi, b)$ w.r.t ϕ can be achieved by solving the gradient flow equation:

$$\frac{\partial\phi}{\partial t} = -\frac{\partial\mathcal{F}}{\partial\phi}. \tag{5.30}$$

We compute the derivative $\frac{\partial\mathcal{F}}{\partial\phi_l}$ with k-phase formulation and $l = 1, \cdots, r$ (the number of level set functions) and re-express Eq. (5.30) as follows:

$$\frac{\partial\phi_l}{\partial t} = -\alpha \sum_{i=1}^{k} \left(\frac{\partial M_i(\phi)}{\partial\phi_l} e_i + 2 \sum_{j=1}^{m} \mathcal{I}_{S_j} \frac{\partial M_i(\phi)}{\partial\phi_l} (\mathcal{I}_{S_j} M_i(\phi) - \frac{h_{ij} a}{\sum_{i=1}^{k} h_{ij}}) \right)$$

$$+ \beta(\nabla^2\phi_l - div(\frac{\nabla\phi_l}{|\nabla\phi_l|})) + \gamma\delta(\phi_l) \, div(g\frac{\nabla\phi_l}{|\nabla\phi_l|}) + \nu g\delta(\phi_l), \tag{5.31}$$

where $\delta(\phi_l)$ is the dirac delta function obtained as the derivative of the Heaviside function.

Then, for fixed ϕ_l, the optimal bias field b that minimizes the energy $\mathcal{F}$ is estimated by:

$$b(\boldsymbol{x}) = \frac{\sum_{i=1}^{k} \int_{\Omega} \frac{I(\boldsymbol{x})\mu_i}{\sigma_i^2} M_i(\phi_l) d(\boldsymbol{x})}{\sum_{i=1}^{k} \int_{\Omega} \frac{\mu_i^2}{\sigma_i^2} M_i(\phi_l) d(\boldsymbol{x})}. \qquad (5.32)$$
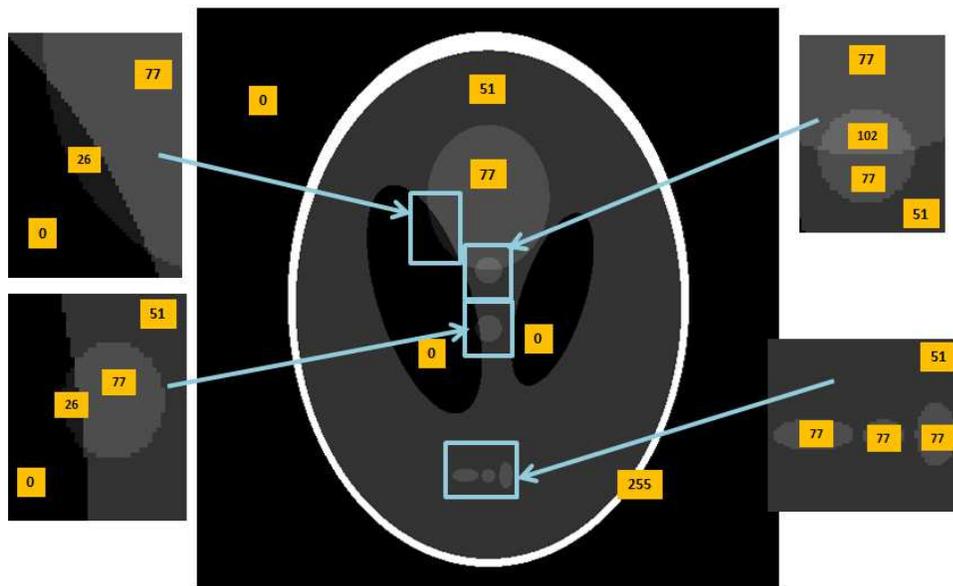
Figure 6. A synthetic gray-scale image with all specified regions (from outside the image $0, 255, 51, 77, 26$, and 102 intensity value).
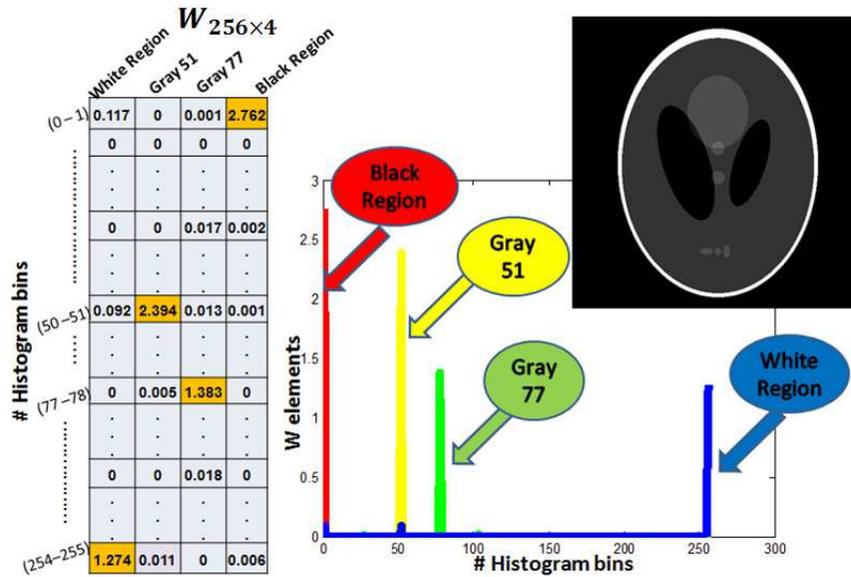
(a) "W" matrix Interpretation.
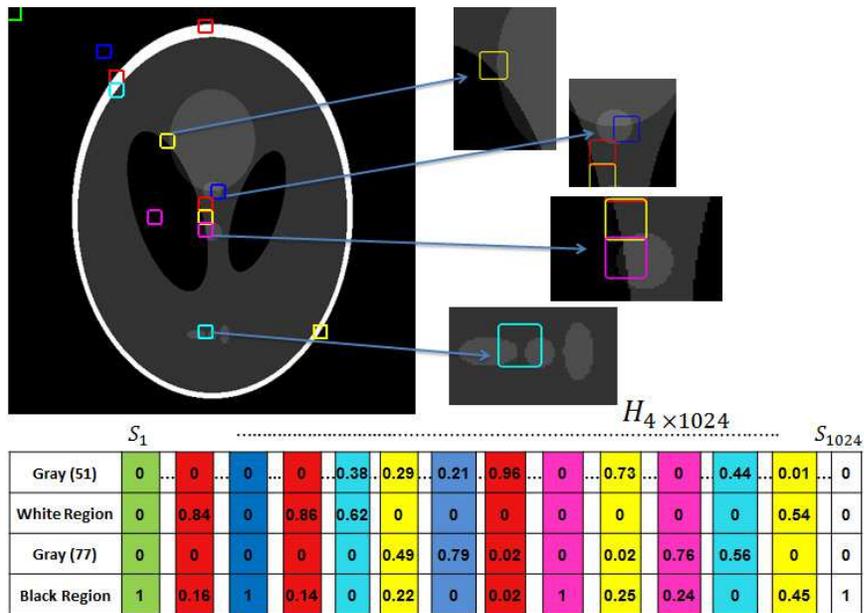
(b) "H" matrix interpretation.

Figure 7. "W" and "H" matrices interpretation for a synthetic gray-scale image with 16×16 block size and the entries of "H" normalized for each column.

$W_{256\times6}$

# Histogram bins	White Region	Gray (51)	Gray (77)	Gray (102)	Gray (26)	Black Region
(0 – 1)	0.025	0	0.001	0	0.035	3.938
	0	0	0	0	0	0
(25 – 26)	0	0	0.004	0	0.693	0
(50 – 51)	0.030	3.408	0.006	0.002	0.001	0
(77 – 78)	0	0.002	2.015	0.178	0.032	0
(102 – 103)	0	0	0.003	0.573	0	0
(254 – 255)	1.950	0.004	0	0	0	0.002

(a) "W" matrix Interpretation.

$H_{6\times1024}$ (S_1 ... S_{1024})

Gray (51)	0	...	0	0.02	...	0	0.23	0.02	0	0	0.08	0.23	0.13	..
White Region	0		0.04	0.98		0	0	0	0	0	0.88	0	0
Gray (77)	0		0	0		0.19	0.49	0.09	0	1	0	0.78	0.87
Gray (102)	0		0			0	0.28	0	0	0	0	0	0
Gray (26)	0		0			0.44	0	0.21	0.35	0	0	0	0
Black Region	1		0.96	0		0.37	0	0.68	0.65	0	0.03	0	0
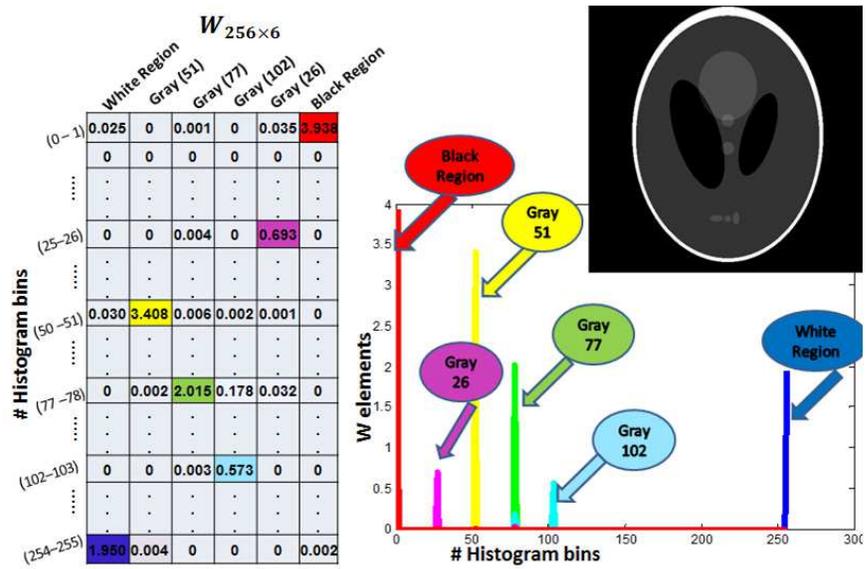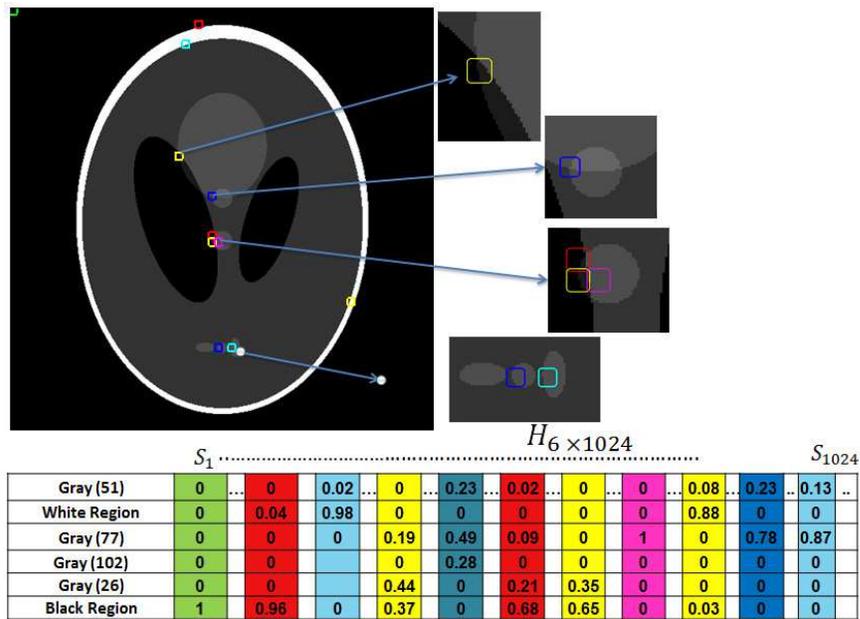
(b) "H" matrix interpretation.

Figure 8. "W" and "H" matrices interpretation for a synthetic gray-scale image with 8×8 block size and the entries of "H" normalized for each column.
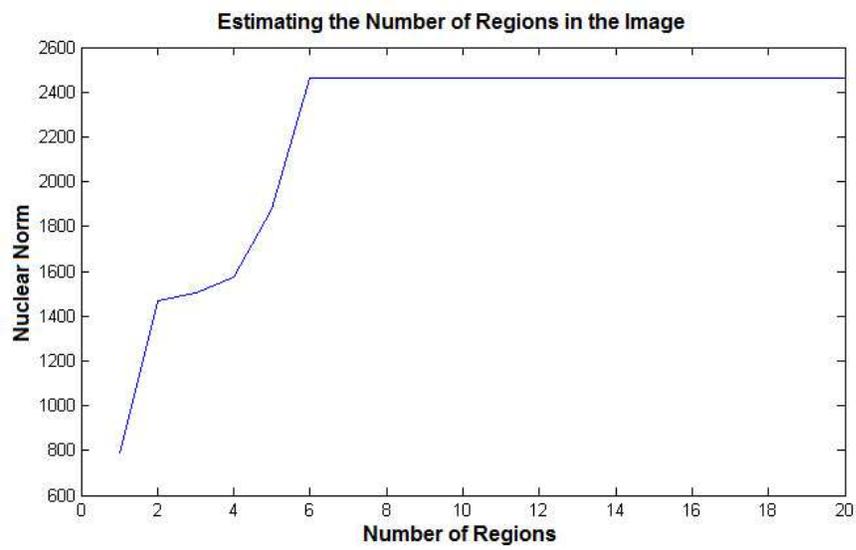
Figure 9. Nuclear norm for the synthetic gray-scale image in Fig. (*6*) which has six regions by increasing k from 1 to 20.

Chapter 6

Simulation Results and Discussion

In the implementation of our proposed PNMF-based level set method, we choose α, β, and γ to be equal to 1 in Eq. (5.31). The smoothed version of the Heaviside function is approximated by $H_\epsilon(x) = 0.5\sin(\arctan(\frac{x}{\epsilon})) + 0.5$, while the dirac delta function, $\delta(x)$, is approximated by $\delta(x) = 0.5\cos(\arctan(\frac{x}{\epsilon}))\frac{\epsilon}{\epsilon^2 + x^2}$. In our simulations, we set $\epsilon = 1$. We automate the initialization of the level set function by using the fuzzy c-means (FCM) algorithm and initiate the level set function as $\phi_o = -4\epsilon(0.5 - B_k)$, where B_k is a binary image obtained from the FCM result. The detailed explanation of FCM used for the initialization is provided in [11]. The weighting parameter ν is defined as $\nu = 2 * (1 - \eta * (2 * B_k + 1))$, for some constants η. We choose the block size to be (8×8) as it is small enough to capture the fine details that we are interested in. Although we choose the block size to be (8×8), we will show that the PNMF interpretation of the matrices W and H carries over at the limit, when the block size is equal to 1. The nuclear norm versus the number of region for block size one in Fig. (*17*) is given as an example. In order to evaluate the performance of the proposed PNMF-based level set method, we first apply it to ten synthetic images whose boundaries are known and used as the ground truth. These images are corrupted with different levels of noise and intensity inhomogeneity. We then compare the performance of the proposed approach with two other state-of-the-art level set models, namely the localized level set model (localized-LSM) [25], and the improved LGDF-LSM model [10]. We study the robustness of our method to the initial conditions. We also study the influence of the weighting parameters α, β, and γ in Eq. (5.28), by choosing different values for each parameter within the range $[0.1, 20]$. We show that the proposed

PNMF-LSM approach carries over at the limit, when the block size is equal to 1 by applying PNMF-LSM with different block sizes on a synthetic gray-scale image used in chapter 5. The only disadvantage of choosing small block size is the computational cost. The quantitative comparison of the segmentation accuracy between our approach and the other level set methods is achieved using three different similarity measures, Jaccard Similarity (JS) [40], Dice coefficient (DC) [3], and root mean square error ($RMSE$). The Jaccard Similarity (JS) is defined as the ratio between the intersection and the union of two regions S_1 and S_2, representing, respectively, the segmented region and the ground truth.

$$\mathcal{J}(S_1, S_2) = \frac{\mid S_1 \cap S_2 \mid}{\mid S_1 \cup S_2 \mid}, \tag{6.1}$$

where $\mid S \mid$ represents the area of region S. The closer the JS to 1 the better the segmentation result. The Dice coefficient (DC) is another metric that measures the spatial overlap between two images or two regions, defined as:

$$\mathcal{D}(S_1, S_2) = \frac{2 \mid S_1 \cap S_2 \mid}{\mid S_1 \cap S_2 \mid + \mid S_1 \cup S_2 \mid}. \tag{6.2}$$

Although Jaccard and Dice coefficients are very similar, the Jaccard similarity is more sensitive when the regions are more similar, while the Dice coefficient gives more weighting to instances where the two images agree [40]. Both of the JS and DC provide values ranging between 0 (no overlap) and 1 (perfect agreement). The root mean square error $RMSE$ is a distance measure that gives the difference between two image regions or image intensities, denoted by R_1 and R_2 as follows,

$$RMSE(R_1(\boldsymbol{x}), R_2(\boldsymbol{x})) = \sqrt{\frac{1}{N} \sum_{\boldsymbol{x} \in \Omega} (R_1(\boldsymbol{x}) - R_2(\boldsymbol{x}))^2}, \tag{6.3}$$

where N is the total number of pixels in the region Ω.

6.1 Performance Evaluation and Comparison

We compare our proposed PNMF-LSM method with the localized-LSM model [25], and the improved LGDF-LSM model [10] using Jaccard similarity, Dice coefficient, and root mean square error. Figure (10) shows three synthetic images with the segmentation results of the three models. We can see that our PNMF-LSM model is able to delineate the boundaries of the objects more accurately than the other two methods, although each image is corrupted with the same level of noise and intensity inhomogeneity. Figure (11) shows the comparison using JS, DC and $RMSE$ values (mentioned in Table (1)) of the three methods on the 10 synthetic images. As shown in Fig. (11) and Table (1), the performance of the proposed PNMF-LSM is more stable with higher JS and DC values than the localized-LSM and the improved LGDF-LSM models. PNMF-LSM also results in a lower mean error rate, while the performance of the other two models changes from one image to another with lower values of JS, DC and higher mean error rate.

6.2 Robustness to Contour Initialization

With the previously mentioned similarity metrics, Jaccard, Dice coefficients and root mean square error, we can quantitatively evaluate the performance of our method starting from different initial conditions. We applied our method to a synthetic image in Fig. (12) with 10 different initializations of the contour. We show three of the 10 initial contours (red contours) and the corresponding segmentation results (green contours) in Fig. (12). In these three different initializations, the initial contour encloses the objects of interest, crosses the objects, and is totally inside one of the objects as displayed in Fig. (12). Starting from these
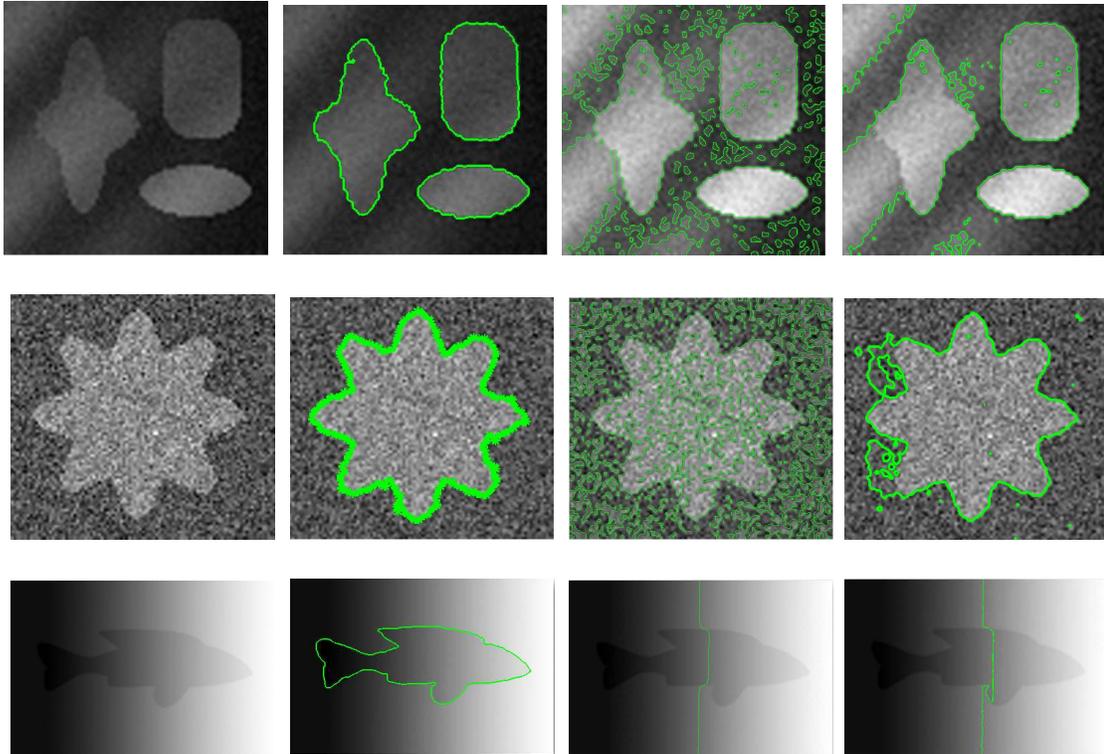
Figure 10. Performance evaluation of the proposed PNMF-LSM, the localized-LSM [25] and the improved LGDF-LSM [10] on three synthetic images corrupted with different level of noise and intensity inhomogeneity. The first column represents the original images to be segmented. The second column shows the segmentation of the proposed PNMF-LSM algorithm. The third and fourth columns show the results of the localized-LSM and the improved LGDF-LSM models, respectively.

48
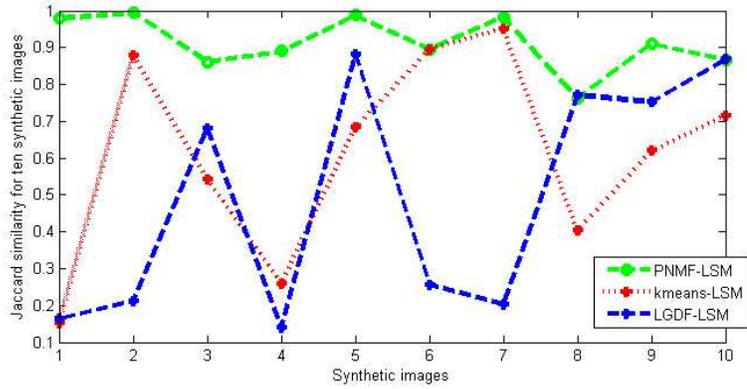
(a) Comparison based on Jaccard similarity.
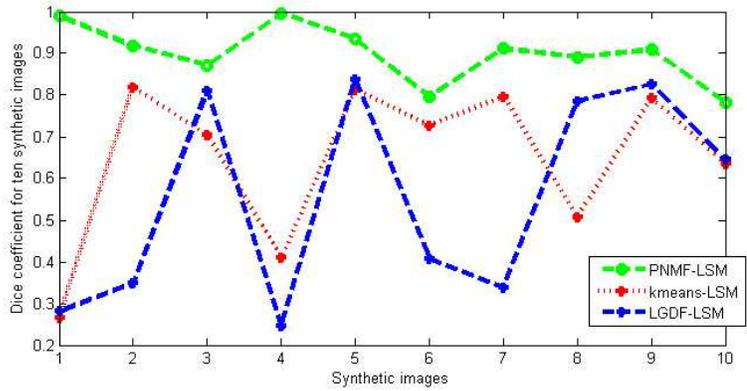
(b) Comparison based on Dice coefficient.
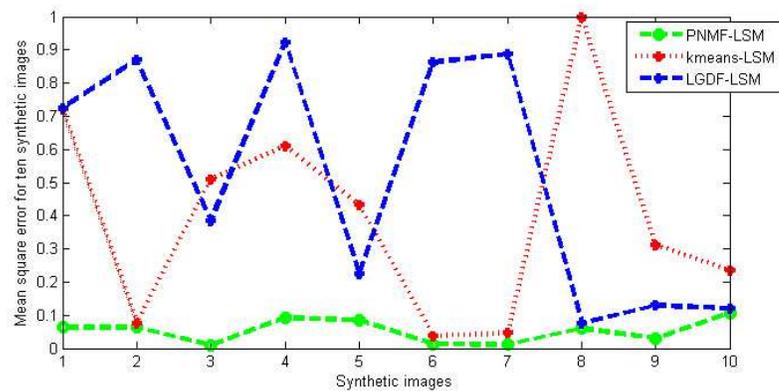
(c) Comparison based on root mean square error.

Figure 11. Comparison based on JS, DC and $RMSE$ values between the three methods, PNMF-LSM, localized-LSM and LGDF-LSM, on 10 synthetic images with different degrees of intensity inhomogeneities and different levels of noise.

Table 1

JS, DC and $RMSE$ similarity measures of the three segmentation models, the proposed PNMF-LSM, the localized-LSM and the improved LGDF-LSM, applied on 10 synthetic images.

Proposed NMF-LSM			Localized-LSM			Improved LGDF-LSM		
JS	DC	$RMSE$	JS	DC	$RMSE$	JS	DC	$RMSE$
0.9788	0.9893	0.0633	0.1534	0.2659	0.7205	0.1634	0.2810	0.7239
0.9943	0.9171	0.0641	0.8780	0.8189	0.0763	0.2121	0.3500	0.8710
0.8600	0.8700	0.0100	0.5422	0.7032	0.5095	0.6813	0.8104	0.3854
0.8904	0.9952	0.0918	0.2586	0.4109	0.6103	0.1397	0.2451	0.9220
0.9880	0.9340	0.0857	0.6844	0.8127	0.4335	0.8815	0.8370	0.2230
0.8948	0.7972	0.0135	0.8948	0.7274	0.0377	0.2562	0.4079	0.8624
0.9840	0.9119	0.0112	0.9502	0.7951	0.0452	0.2035	0.3382	0.8879
0.7621	0.8910	0.0593	0.4035	0.5069	0.9976	0.7706	0.7851	0.0761
0.9100	0.9095	0.0303	0.6200	0.7920	0.3100	0.7527	0.8263	0.1292
0.8663	0.7813	0.1967	0.7161	0.6346	0.2358	0.8685	0.6465	0.1212

initial contours, the corresponding segmentation results are almost the same, all accurately capturing the objects' boundaries. The segmentation accuracy is quantitatively assessed in terms of the Jaccard similarity, the Dice coefficient and the root mean square error. The Jaccard and Dice coefficients of these results are all between 0.78 and 0.97 pixel, while the root mean square error is between 0.03 and 0.1 as shown in Fig. (*13*). These experiments demonstrate the robustness of our PNMF-LSM model to contour initialization.

6.3 Stable Performance for Different Weighting Parameters

We compare our PNMF-LSM model with the localized-LSM and the improved LGDF-LSM models for different weighting parameters α, β and γ in Eqs. (5.28) and (5.31). Fig-

ure (*14*) shows the box plots of the JS, DC and $RMSE$ values for each method (PNMF-LSM, localized-LSM and improved LGDF-LSM). From the box plots of the JS, DC and $RMSE$ values, it is clear that the PNMF-LSM has better and more stable performance in terms of segmentation accuracy and robustness. We notice that the boxes shown in the box plots of the PNMF-LSM are relatively shorter with higher JS, DC values and lower error rate for different values of α, β and γ. Table (2) shows different values of JS, DC and $RMSE$ obtained from our PNMF-LSM model and the other two level set models for different values of α, β and γ.

At the same time, our model is much more computationally efficient than the localized-LSM and the improved LGDF-LSM models. This can be seen from the CPU times of the three models in Fig. (*15*) and Table (3). In this experiment, our PNMF-LSM model is remarkably faster than the other two models for different values of α, β and γ. The CPU times were recorded from Matlab programs on a Asus K53E laptop with Intel(R) Core(TM)i5-2450M CPU, 2.50 GHz, 8 GB RAM, with Matlab R2013a on Windows 7. Figure (*15*) shows the convergence time of the three models for different values of α, β and γ, by using the box plots. It can be seen from Fig. (*15*) that the boxes of our model are relatively shorter with lower values of the convergence time than the localized-LSM and the improved LGDF-LSM models.

6.4 PNMF-LSM Evaluation for Small Block Sizes

We apply the proposed PNMF-LSM approach on the synthetic gray-scale image in chapter 5 with different block sizes (16×16), (8×8), (4×4), (2×2) and (1×1). The segmentation accuracy is quantitatively assessed in terms of the Jaccard similarity, the Dice coefficient

and the root mean square error. The Jaccard and Dice coefficients of these results are all between 0.74 and 0.98 pixel, while the root mean square error is between 0.001 and 0.06 as shown in Fig. (*16*) and Table (4). This experiment demonstrates that decreasing the block size to be very small (less than (8×8) in this case) will not affect the segmentation accuracy of the proposed approach. It also demonstrates that the proposed PNMF-LSM algorithm carries over at the limit, when the block size is equal to 1. Figure (*17*) shows the nuclear norm versus the number of region for block size one for the synthetic gray-scale image used in chapter 5. We notice that the nuclear norm stabilizes at $k = 6$ for block size 1×1, which is the true number of regions in the image. The disadvantage of decreasing the block size to be very small is the computational cost. This can be seen from the CPU times of the proposed PNMF-LSM with different block sizes in Table (4).
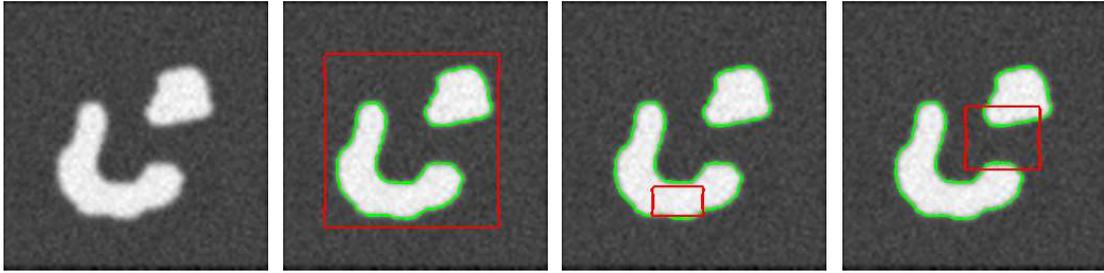
Figure 12. Robustness of the proposed PNMF-LSM segmentation to contour initializations. The initial contours are represented by the red contours and the corresponding segmentation results are represented by the green contours.
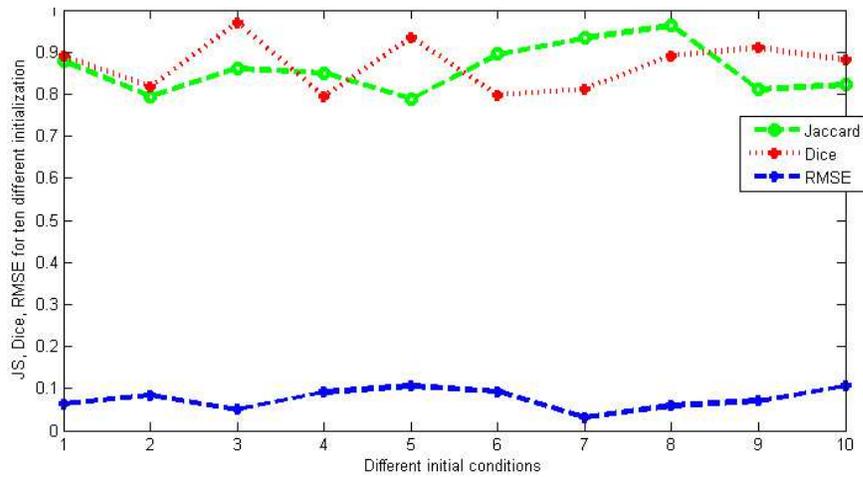
Figure 13. Segmentation accuracy of the proposed PNMF-LSM approach for different initial contours as measured by the JS, DC and $RMSE$.
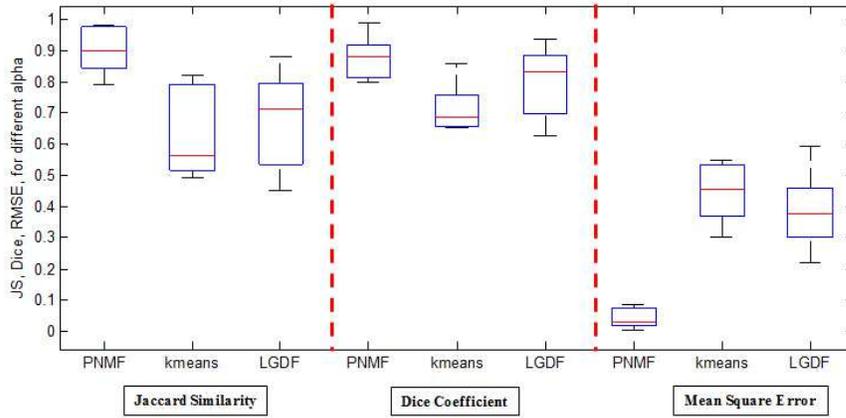
(a) JS, DC and $RMSE$ for different α.
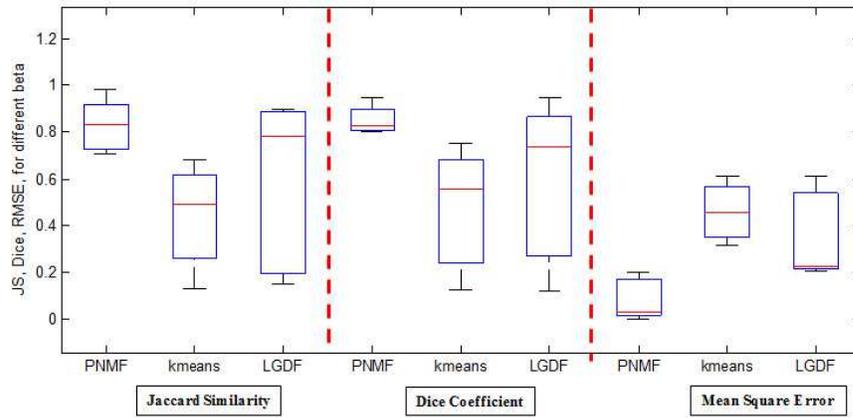
(b) JS, DC and $RMSE$ for different β.
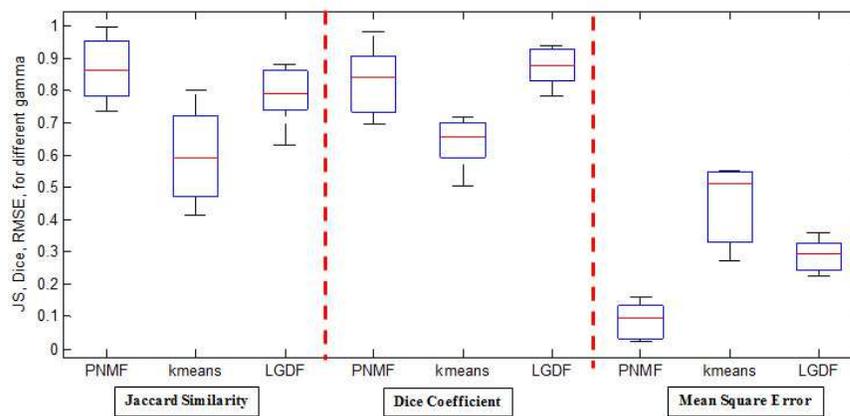
(c) JS, DC and $RMSE$ for different γ.

Figure 14. The box plots of the JS, DC and $RMSE$ values for the star object in the synthetic image obtained from the proposed PNMF-LSM, the localized-LSM and the improved LGDF-LSM for different values of the parameters α, β and γ.

54
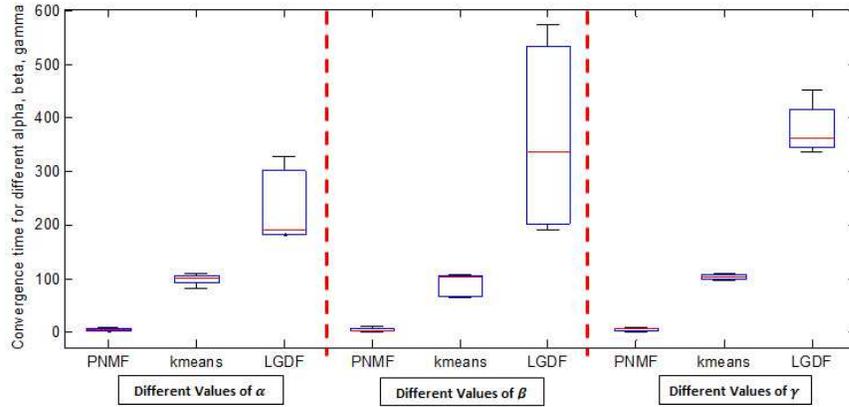
Figure 15. The box plots of convergence times of the three models, the proposed PNMF-LSM, the localized-LSM and the improved LGDF-LSM, for different value of the parameters α, β and γ.
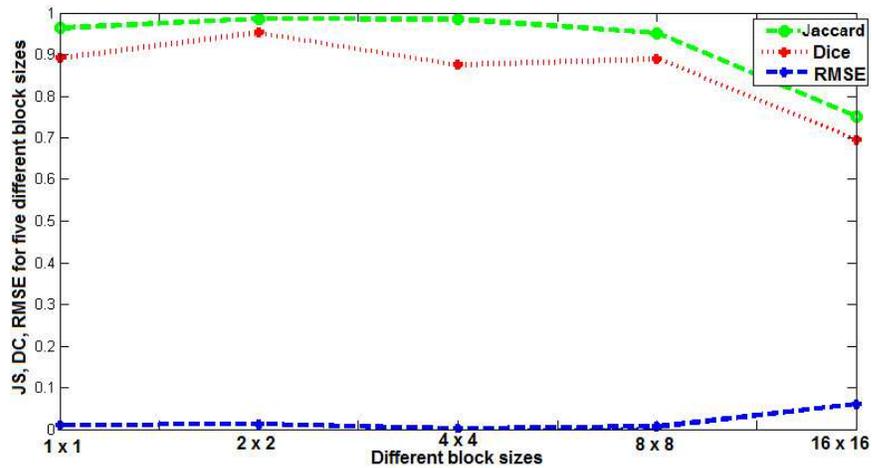
Figure 16. PNMF-LSM performance evaluation for different block sizes using the JS, DC and $RMSE$.

Table 2

JS, DC and RMSE similarity measures of the three segmentation models, the proposed PNMF-LSM, the localized-LSM and the improved LGDF-LSM, for different values of α, β and γ.

	Proposed PNMF-LSM			**Localized-LSM**			**Improved LGDF-LSM**		
α	JS	DC	$RMSE$	JS	DC	$RMSE$	JS	DC	$RMSE$
0.1	0.7902	0.8191	0.0303	0.7824	0.6509	0.4544	0.7125	0.8321	0.3760
0.5	0.9801	0.8800	0.0243	0.8201	0.6578	0.3919	0.7677	0.8686	0.3286
1	0.8990	0.7993	0.0010	0.4927	0.8602	0.5504	0.8815	0.9370	0.2230
10	0.9753	0.9862	0.0862	0.5235	0.6872	0.5314	0.5620	0.7196	0.4133
20	0.8622	0.8991	0.0712	0.5637	0.7210	0.3022	0.4531	0.6236	0.5944
β	JS	DC	$RMSE$	JS	DC	$RMSE$	JS	DC	$RMSE$
0.1	0.8330	0.8091	0.0303	0.6841	0.7524	0.3586	0.7831	0.8379	0.2211
0.5	0.7090	0.7992	0.0010	0.4875	0.6555	0.4548	0.8958	0.9451	0.2075
1	0.9794	0.8791	0.0200	0.5927	0.5602	0.5504	0.8815	0.7370	0.2230
10	0.7353	0.8262	0.1584	0.300	0.2800	0.6139	0.2100	0.3200	0.6139
20	0.8960	0.9452	0.2011	0.1320	0.1260	0.3139	0.1500	0.1200	0.5139
γ	JS	DC	$RMSE$	JS	DC	$RMSE$	JS	DC	$RMSE$
0.1	0.7971	0.6991	0.0950	0.5916	0.6591	0.5511	0.6329	0.8459	0.3576
0.5	0.7391	0.8791	0.1231	0.4920	0.7195	0.5108	0.8569	0.9230	0.2471
1	0.9364	0.7472	0.1581	0.4127	0.6202	0.2704	0.8815	0.9370	0.2230
10	0.8621	0.8397	0.0237	0.6965	0.6935	0.3480	0.7786	0.8755	0.3172
20	0.9971	0.9814	0.0348	0.7992	0.5060	0.5463	0.7883	0.7816	0.2949

Table 3

Convergence time of the three segmentation models, the proposed PNMF-LSM, the localized-LSM and the improved LGDF-LSM, for different values of the parameters α, β and γ (measured in seconds).

α	**Proposed PNMF-LSM**	**Localized-LSM**	**Improved LGDF-LSM**
0.1	2.246	81.88	184.46
0.5	6.997	102.20	183.44
1	4.165	103.74	192.13
10	10.794	110.10	293.77
20	4.887	97.89	327.43
β	**Proposed PNMF-LSM**	**Localized-LSM**	**Improved LGDF-LSM**
0.1	10.981	108.31	205.77
0.5	6.965	105.56	192.71
1	1.825	103.74	336.29
10	2.950	68.72	574.63
20	3.840	64.53	521.36
γ	**Proposed PNMF-LSM**	**Localized-LSM**	**Improved LGDF-LSM**
0.1	6.809	101.46	452.57
0.5	7.950	110.07	402.89
1	1.872	103.74	336.29
10	3.856	107.37	363.47
20	10.809	98.44	348.434
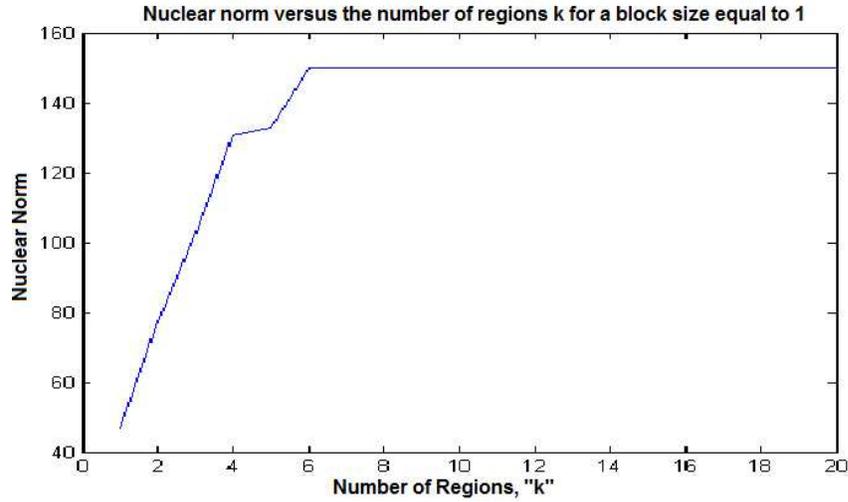
Figure 17. Nuclear norm with a block size one for the synthetic gray-scale image in Fig. (*6*) which has six regions by increasing k from 1 to 20.

Table 4

JS, DC, $RMSE$ and CPU time (in seconds) of the proposed PNMF-LSM approach with different block sizes $(16 \times 16), (8 \times 8), (4 \times 4), (2 \times 2)$ and (1×1).

	Proposed PNMF-LSM			
Block size	JS	DC	$RMSE$	CPU Time
1×1 pixel	0.9635	0.8911	0.0101	92.896
2×2 pixel	0.9855	0.9525	0.0120	46.235
4×4 pixel	0.9845	0.8756	0.0010	16.589
8×8 pixel	0.9511	0.8892	0.0060	4.165
16×16 pixel	0.7512	0.7456	0.0600	1.825

Chapter 7

Application to Real Brain MRI Images

In this chapter, we focus on the application of the proposed PNMF-LSM algorithm on real brain MRI images with and without Glioblastoma. The aim is to segment the different brain and tumor structures. We first apply a preprocessing step on the brain MRI images: histogram equalization and morphological operations. Histogram equalization improves the contrast of the image by spreading out the most frequent intensity values. We subsequently remove the non-brain structures using morphological operations. Specifically, we used *thresholding* to remove the background, *erosion* to shrink the brain and skull, *opening* to remove the small non-brain structures, *labeling* to isolate the brain from the skull and non-brain structures and *dilation* to recover the exact boundaries of the brain. Figures (*18*), (*19*), (*20*) show the segmentation result of our proposed PNMF-LSM approach on different MRI slices of the brain. We notice from the figures that our model is able to delineate the different brain structures: gray matter, white matter, CSF, edema, and tumor with the necrosis inside. Although we applied histogram equalization to enhance the contrast in the MRI images, they are still corrupted with intensity inhomogeneity, which the proposed PNMF-LSM approach is able to handle. We can see in Figs. (*18*), (*19*), (*20*), that although the intensities of the gray matter, the white matter and the edema (flair) are very close to each other and their histograms overlap, PNMF-LSM is able to separate them using an 8×8 block size, which is small enough to capture the fine details in the image. The PNMF retrieves the histogram of each region (or brain structure, see the right image in the second row). We subsequently compute the means and variances of each brain structure and use them to delineate the boundaries of these structures in the level-set framework, as described

in Chapter 5. We show also the bias field in the MRI images (right image in the first row), which is estimated through the level set formulation. In Fig. (21), we apply our proposed PNMF-LSM on MRI images of normal brain without tumor. It is seen from Fig. (21) that PNMF-LSM is able to separate the gray matter, the white matter and the CSF. The binary representation of each brain structure shows the exact boundaries of these structures.

7.1 Robustness to Noise

In order to show the robustness of our method to noise introduced in the brain MRI images, we add two types of noise: i) blurring by Gaussian noise with standard deviation equal to (2) and ii) salt and pepper noise with density (15%) . Then we apply our proposed PNMF-LSM algorithm, the localized-LSM and the improved LGDF-LSM methods on the same MRI images with and without noise. The segmentation result of the PNMF-LSM on the brain MRI image without noise is shown in Fig. (22). Figure (22) shows the contours with the binary representation of each brain structure: gray matter, white matter, tumor with the necroses, edema (flair) and CSF with background. It also shows the bias field and the histogram of each brain structure obtained from the matrix W in the PNMF factorization. In Fig. (23), the same MRI image of Fig. (22) is blurred by Gaussian noise with standard deviation equal to (2). We notice that the blurring did not change the boundaries of each segmented brain structure. Similar segmentation performance in Fig. (24) was obtained when the image was corrupted with salt and pepper noise. The binary representations of each brain structure in Figs. (22), (23), (24) are visually indistinguishable. The robustness to noise of the PNMF-LSM approach can be explained by the fact that the clustering model relies on the histogram rather than the image absolute intensity values. We also applied

the localized-LSM approach [25] on the same MRI images with and without noise. Figure (25) shows the segmentation result of the localized-LSM model on the same MRI image in Fig. (22). We observe that the localized-LSM is not able to separate the tumor region from the white matter and the flair (or edema) region from the gray matter. Moreover, the localized LSM boundary of each brain structure (each region) is not accurately delineated. Finally, we applied the improved LGDF-LSM model [10] on the same MRI images with and without noise. We notice in Fig. (28) that the improved LGDF-LSM model is also unable to separate the brain regions accurately. For instance, we can see in Fig. (28) that part of the flair region is merged with the gray matter and the other part is merged with the CSF and background region. Also, parts of the white matter region appear in the CSF. The improved LGDF-LSM model is also not able to capture the "roots" of the tumor. When the image was blurred by gaussian noise in Fig. (29), we noticed that the white matter region is more obvious but merged with the tumor. The edema (flair) region still appears with the CSF and the background while part of it merged with the gray matter. We obtained similar results when we added salt and pepper noise and applied the improved LGDF-LSM model. The unreliable segmentation results of the localized-LSM and the improved LGDF-LSM models are caused by the overlapping histograms of the gray matter, white matter, edema, CSF and the tumor, and especially the gray matter with the edema (flair), and the white matter with the tumor. These overlapping histograms made the models misclassify the regions accurately. The proposed PNMF-LSM approach is able to handle this problem because it considers local histograms and resizable block sizes that can fit into the small details.
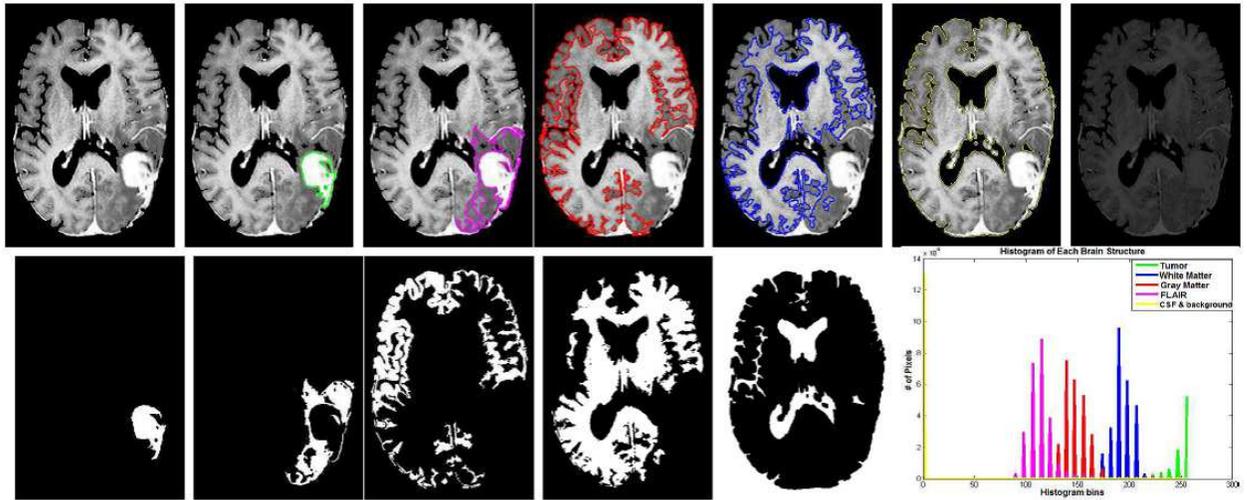
Figure 18. Segmentation of a brain MRI image with Glioblastoma using the proposed PNMF-LSM approach. First row from the left: original image, tumor, flair (or edema), gray matter, white matter, CSF with the background and the estimated bias field. Second row from the left: binary representations of the tumor, flair (edema), gray matter, white matter and CSF with the background; and the histogram of each brain structure obtained from the factor matrix W.
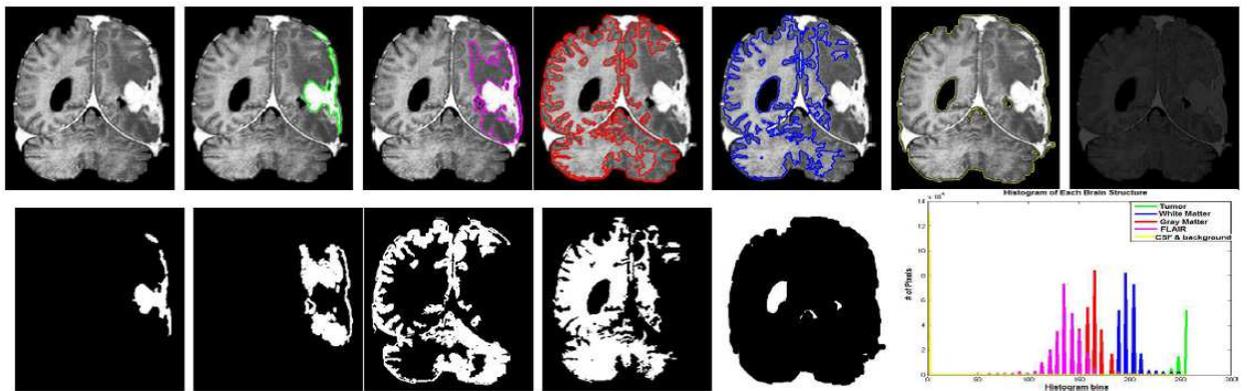
Figure 19. Segmentation of a brain MRI image with Glioblastoma using the proposed PNMF-LSM approach. First row from the left: original image, tumor, flair (edema), gray matter, white matter, CSF with the background and the estimated bias field. Second row from the left: binary representations of the tumor, flair (edema), gray matter, white matter and CSF with the background; and the histogram of each brain structure obtained from the matrix W.
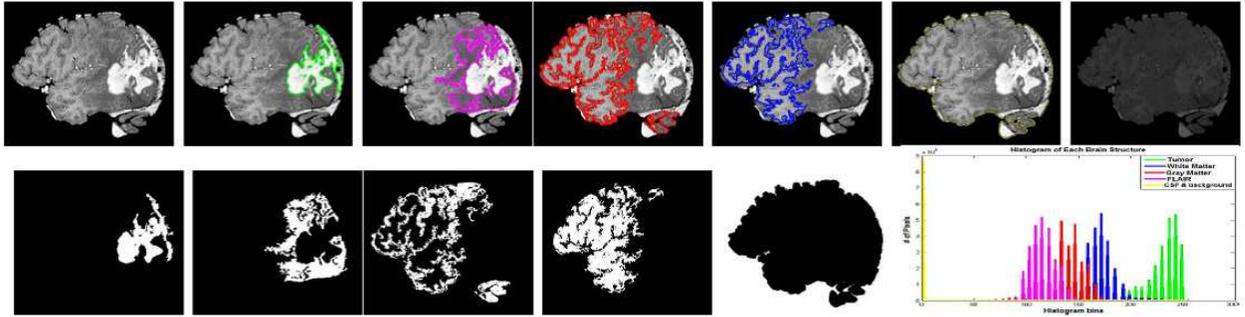
Figure 20. Segmentation of a brain MRI image with Glioblastoma using the proposed PNMF-LSM approach. First row from the left: original image, tumor, flair (edema), gray matter, white matter, CSF with the background and the estimated bias field. Second row from the left: binary representations of the tumor, flair (edema), gray matter, white matter and CSF with the background; and the histogram of each brain structure obtained from the matrix W.
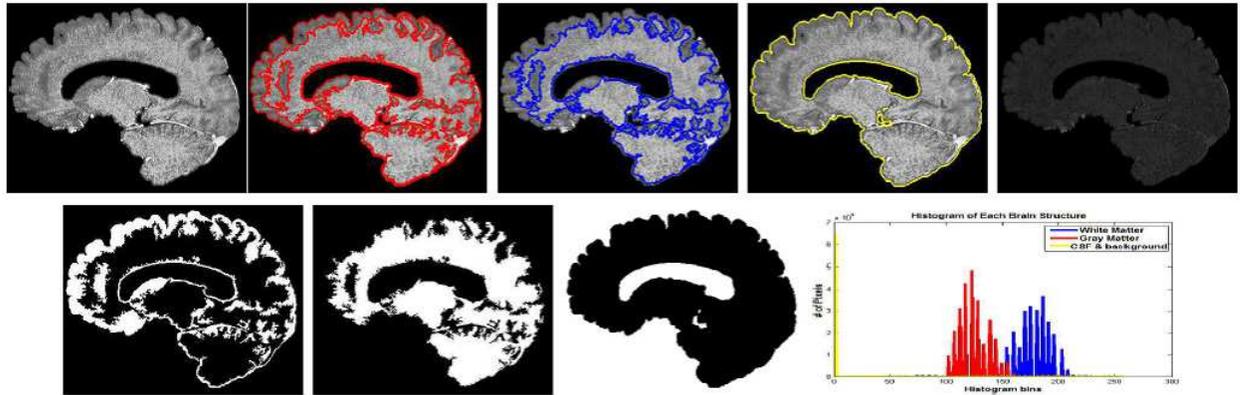
Figure 21. Segmentation of a normal brain MRI image using the proposed PNMF-LSM approach. First row from the left: original image, gray matter, white matter, CSF with the background and the estimated bias field. Second row from the left: binary representations of the gray matter, white matter and CSF with the background; and the histogram of each brain structure obtained from the matrix W.
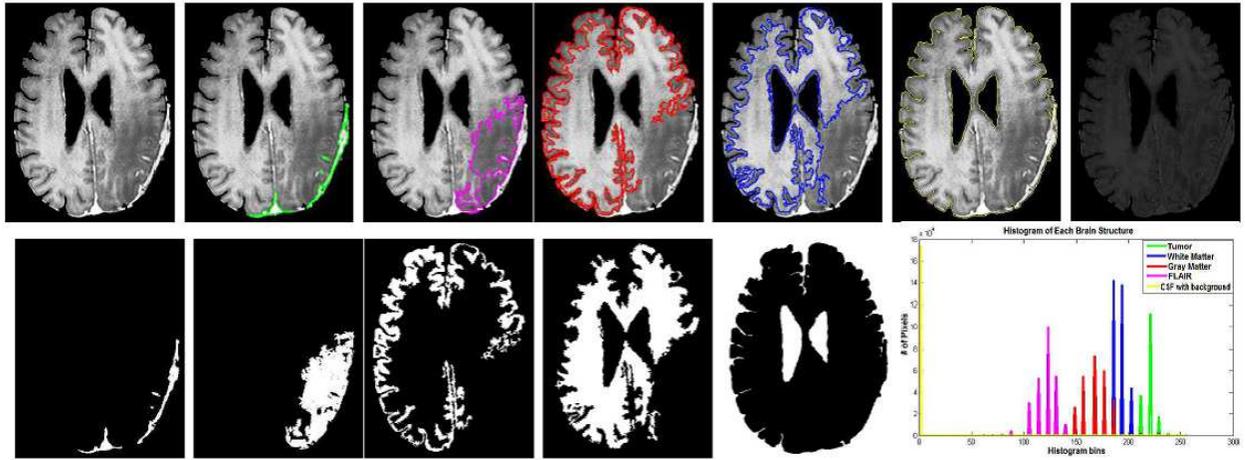
Figure 22. Segmentation of a brain MRI image with Glioblastoma using the proposed PNMF-LSM approach. First row from the left: original image, tumor, flair (edema), gray matter, white matter, CSF with the background and the estimated bias field. Second row from the left: binary representations of the tumor, flair (edema), gray matter, white matter and CSF with the background; and the histogram of each brain structure obtained from the matrix W.
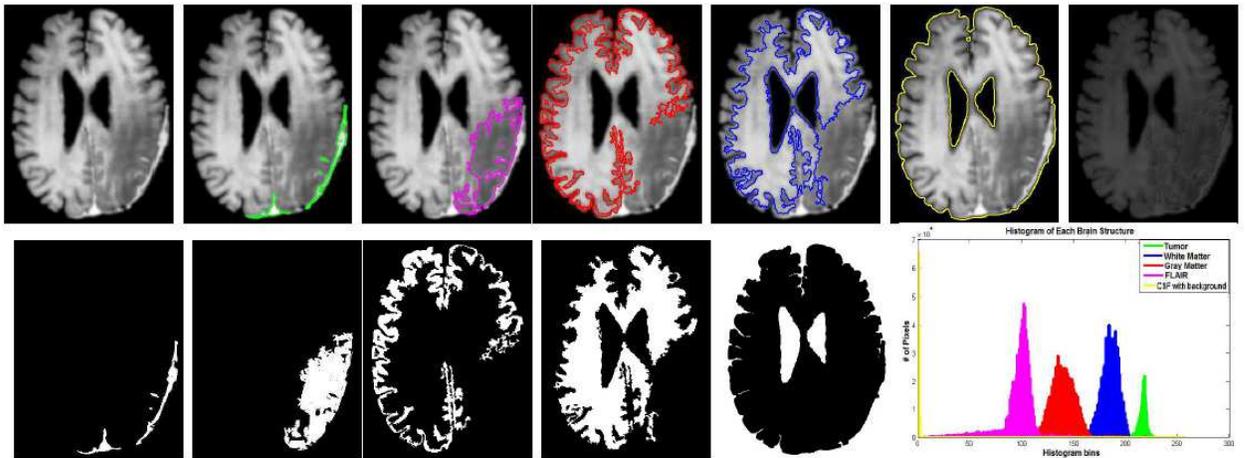
Figure 23. PNMF-LSM segmentation of a brain MRI image with Glioblastoma blurred by Gaussian noise with standard deviation (2). First row from the left: the original image, tumor, flair (edema), gray matter, white matter, CSF with the background and the estimated bias field. Second row from the left: binary representations of the tumor, flair (edema), gray matter, white matter and CSF with the background; and the histogram of each brain structure obtained from the matrix W.
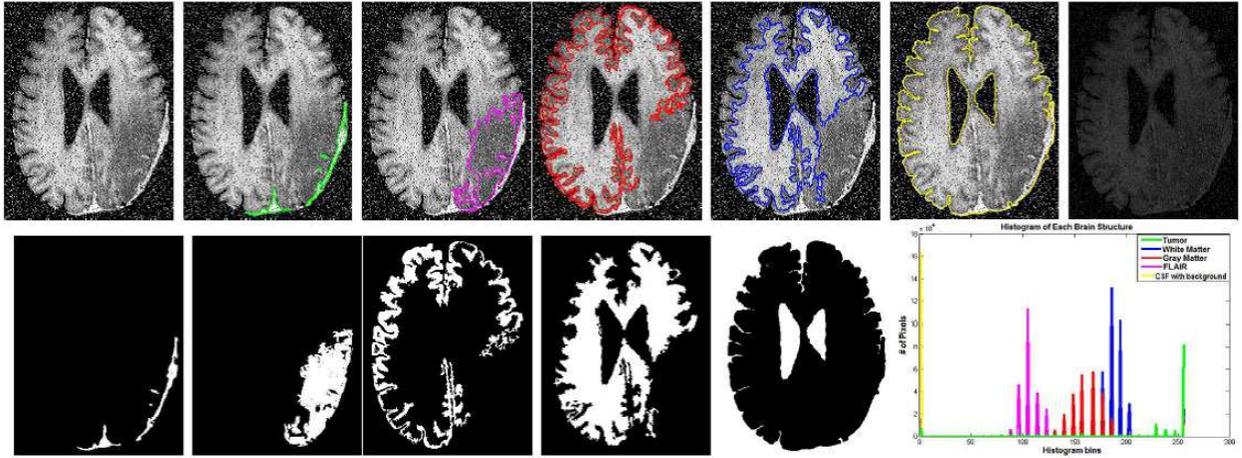
Figure 24. PNMF-LSM segmentation of a brain MRI image with Glioblastoma corrupted by salt and pepper noise. First row from the left: the original image, tumor, flair (edema), gray matter, white matter, CSF with the background and the estimated bias field. Second row from the left: binary representations of the tumor, flair (edema), gray matter, white matter and CSF with the background; and the histogram of each brain structure obtained from the matrix W.
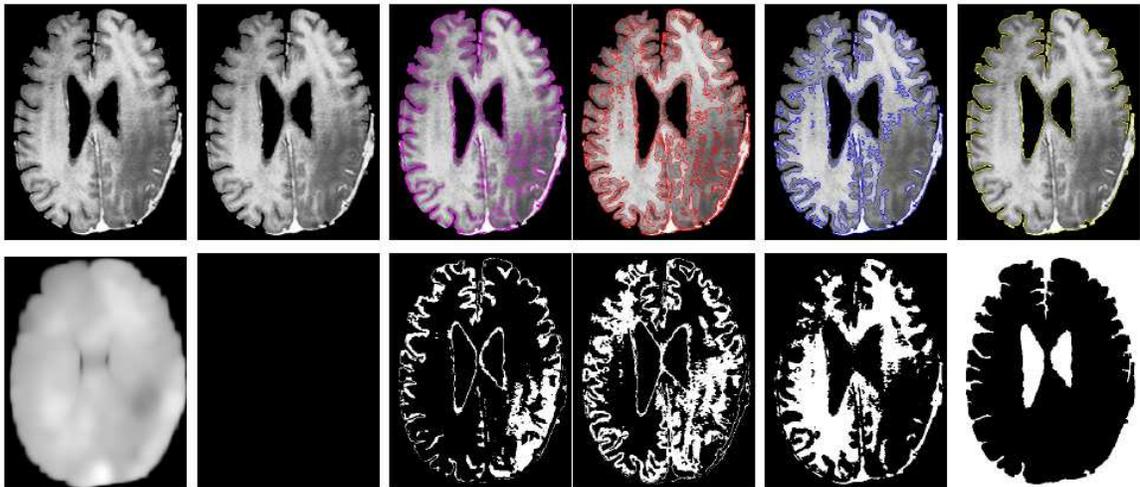
Figure 25. Localized-LSM segmentation of a brain MRI image with Glioblastoma. First row from the left: the original image, tumor, flair (edema), gray matter, white matter and CSF with the background. Second row from the left: the estimated bias field and the binary representations of the tumor, flair (edema), gray matter, white matter and CSF with the background.
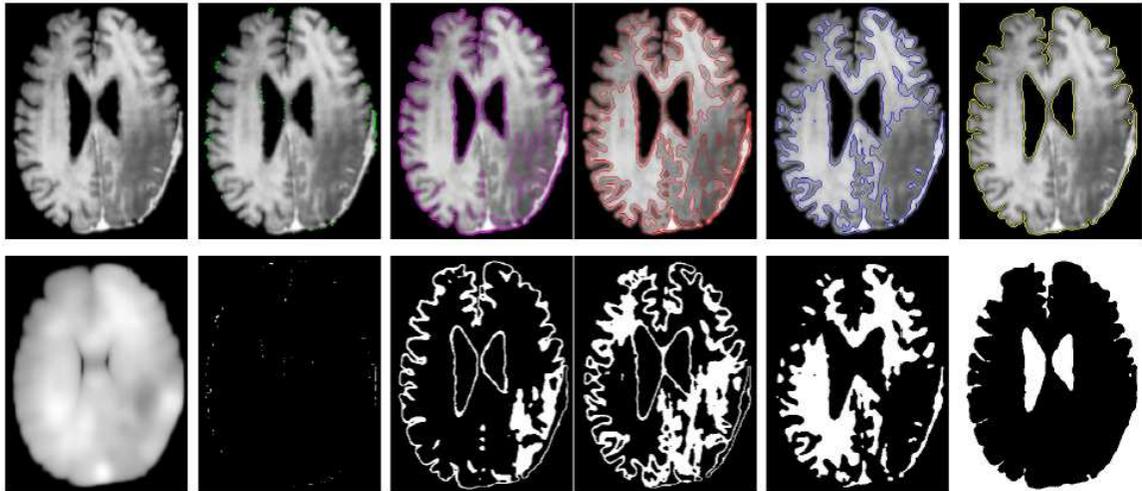
Figure 26. Localized-LSM segmentation of a brain MRI image with Glioblastoma blurred by Gaussian noise. First row from the left: the original image, tumor, flair (edema), gray matter, white matter and CSF with the background. Second row from the left: the estimated bias field and the binary representations of the tumor, flair (edema), gray matter, white matter and CSF with the background.
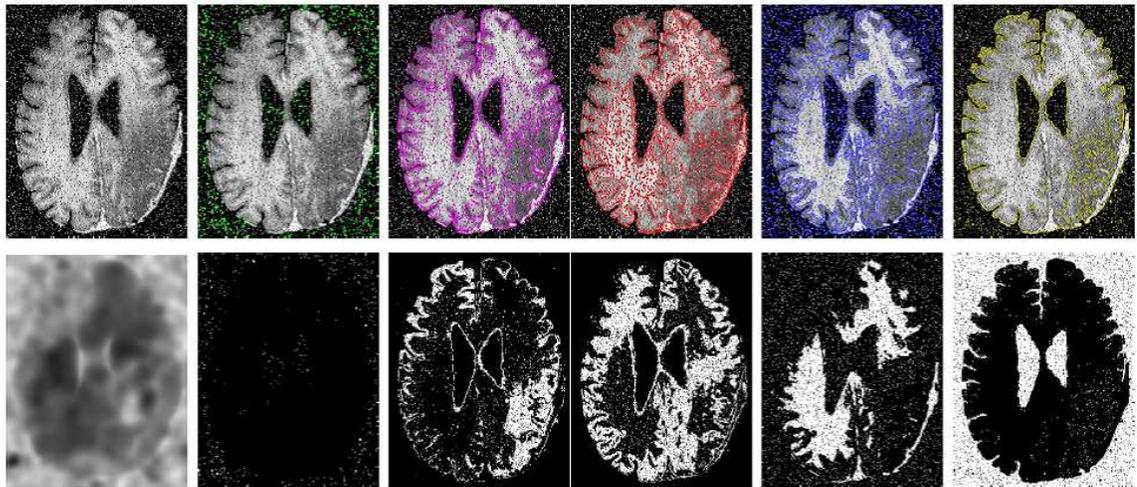
Figure 27. Localized-LSM segmentation of a brain MRI image with Glioblastoma corrupted by salt and pepper noise. First row from the left: the original image, tumor, flair (edema), gray matter, white matter and CSF with the background. Second row from the left: the estimated bias field and the binary representations of the tumor, flair (edema), gray matter, white matter and CSF with the background.
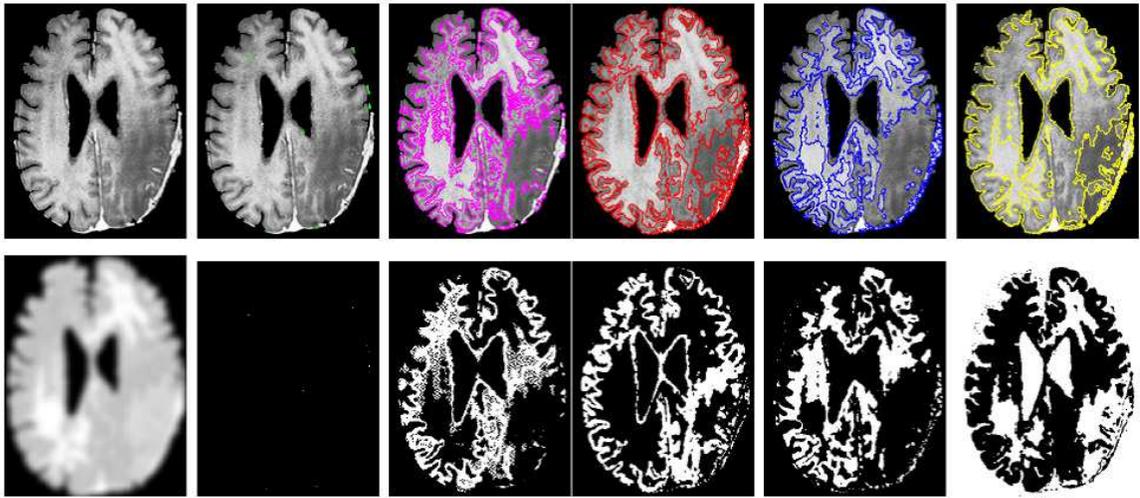
Figure 28. Improved LGDF-LSM segmentation of a brain MRI image with Glioblastoma. First row from the left: the original image, tumor, flair (edema), gray matter, white matter and CSF with the background. Second row from the left: the estimated bias field and the binary representations of the tumor, flair (edema), gray matter, white matter and CSF with the background.
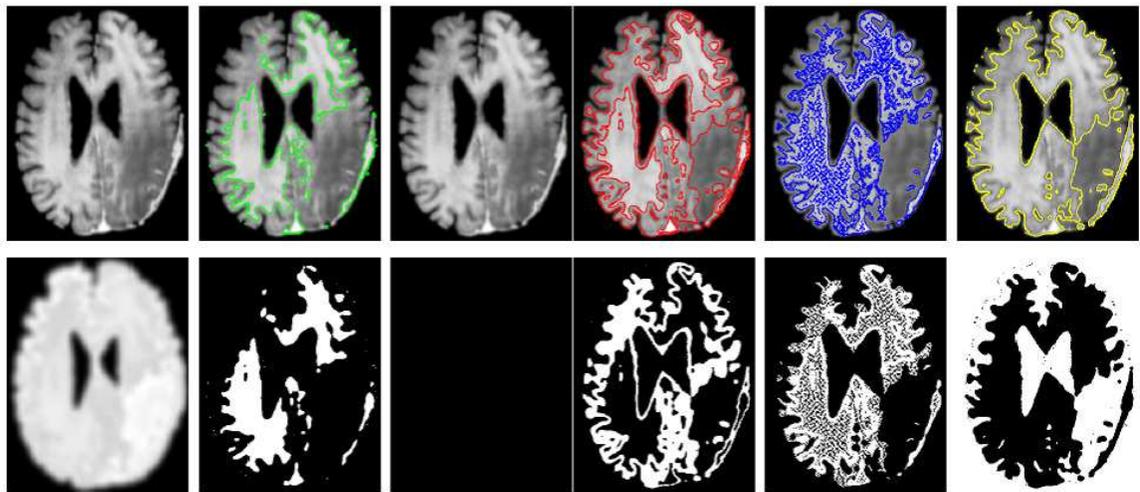
Figure 29. Improved LGDF-LSM segmentation of a brain MRI image with Glioblastoma blurred by Gaussian noise. First row from the left: the original image, tumor, flair (edema), gray matter, white matter and CSF with the background. Second row from the left: the estimated bias field and the binary representations of the tumor, flair (edema), gray matter, white matter and CSF with the background.
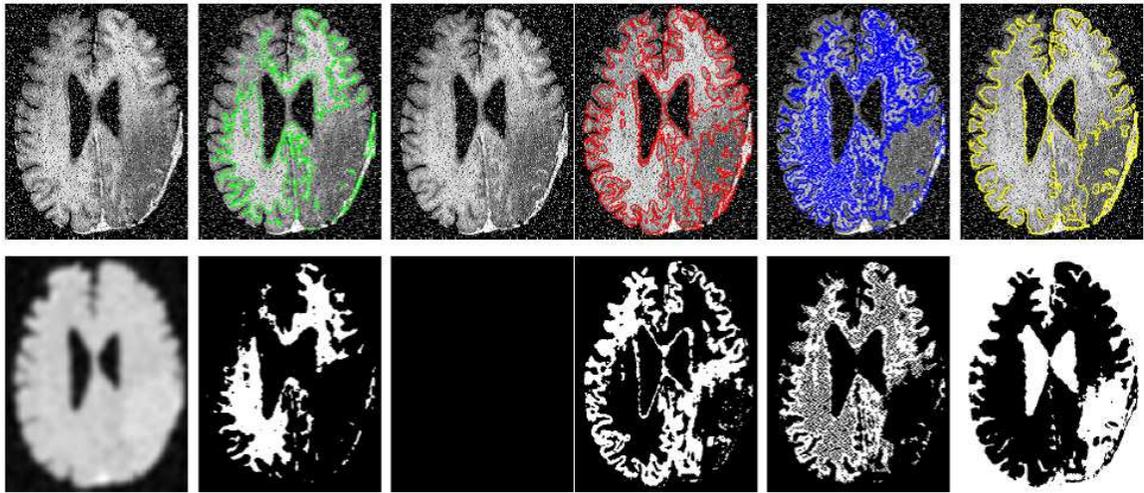
Figure 30. Improved LGDF-LSM segmentation of a brain MRI image with Glioblastoma corrupted by salt and pepper noise. First row from the left: the original image, tumor, flair (edema), gray matter, white matter and CSF with the background. Second row from the left: the estimated bias field and the binary representations of the tumor, flair (edema), gray matter, white matter and CSF with the background.

Chapter 8

Summary and Conclusion

Medical image segmentation is one of the most important and challenging tasks in medical image analysis. In particular, brain MRI segmentation is a difficult task, due to the complexity of the brain anatomical structures and the intensity inhomogeneity that corrupts the quality of MRI images. Great efforts have been made in this field in order to achieve automatic accurate segmentation results. In this thesis, we proposed a new deformable model for image segmentation based on variational level set formulation and probabilistic non-negative matrix factorization, termed PNMF-LSM. The main advantages of the proposed PNMF-LSM approach over the state-of-the-art, as well as the contributions of this thesis are:

- Relying on the histogram data of the image for clustering rather than the pixel intensity values; thus making the algorithm robust to additional noise and outliers.

- Estimating the number of regions/clusters in the image based on the nuclear norm of the PNMF factors (proof provided).

- Providing useful interpretation of the NMF as an image clustering and decomposition tool.

- Deriving two data terms based on the PNMF factors.

- No other method in the LSM literature introduced a spatial term in addition to the intensity-model terms.

- No additional nuisance or spurious model parameters are simultaneously estimated

with the bias field and the level set functions. This increases the estimation accuracy

of the main parameters; thus leading to a higher segmentation accuracy.

References

[1] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T.Moriarty. Amodified fuzzy c-means algorithm for bias field estimation and segmentation of mri data. *IEEE Transactions on Medical Imaging*, 21:193–199, 2002.

[2] R. Albright, J. Cox, D. Duling, A. Langville, and C. Meyer. Algorithms, initializations, and convergence for the nonnegative matrix factorization. *NCSU Technical Report Math*, pages 1–18, 2006.

[3] K. O. Babalola, B. Patenaude, P. Aljabar, J. Schnabel, D. Kennedy, W. Crum, S. Smith, T. F. Cootes, M. Jenkinson, and D. Rueckert. Comparison and evaluation of segmentation techniques for subcortical structures in brain mri. *Medical Image Computing and Computer-Assisted Intervention*, 5241:409–416, 2008.

[4] B. Bayar, N. Bouaynaya, and R. Shterenberg. Probabilistic non-negative matrix factorization: Theory and application to microarray data analysis. *Journal of Bioinformatics and Computational Biology*, 12:25, 2014.

[5] J. P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov. Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences*, 101:4164 – 4169, 2004.

[6] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

[7] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10:266–277, 2001.

[8] F. Charles. *An introduction to the calculus of variations*. Courier Corporation, 1987.

[9] S. Chen and D. Zhang. Robust image segmentation using fcm with spatial constraints based on new kernel-induced distance measure. *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics*, 34:1907–1916, 2004.

[10] Y. Chen, J. Zhang, A. Mishra, and J. Yang. Image segmentation and bias correction via an improved level set method. *Neurocomputing*, 74(17):3520 – 3530, 2011.

[11] K. S. Chuang, H. L. Tzeng, S. Chen, J. Wu, and T. J. Chen. Fuzzy c-means clustering with spatial information for image segmentation. *Computerized Medical Imaging and Graphics*, 30:9–15, 2006.

[12] L. D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, 1993.

[13] D. Cremers, F. Tischhuser, J. Weickert, and C. Schnrr. Diffusion snakes: introducing statistical shape knowledge into the mumford-shah functional. *Journal of Computer Vision*, 50:295–313, 2002.

[14] I. Despotovic, B. Goossens, E. Vansteenkiste, and W. Philips. T1- and t2-weighted spatially constrained fuzzy c-means clustering for brain mri segmentation. *in Medical Imaging: Image Processing, Proceedings of SPIE, p. 9, San Diego, Calif, USA*, 2010.

[15] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons, New York, NY,USA, 2nd edition, 2001.

[16] R. C. Gonzalez and R. E. Woods. *Digital Image Processing. 3nd. ed*. New Jersey: Pearson Prentice Hall, 2008.

[17] M. Gudmundsson and M. R. Kabuka. Edge detection in medical images using a genetic algorithm. *IEEE Transactions on Medical Imaging*, 17(1):469–474, 1998.

[18] T. Hancewicz and J. Wang. Discriminant image resolution: A novel multivariate image analysis method utilizing a spatial classification constraint in addition to bilinear nonnegativity. *Chemometrics and Intelligent Laboratory Systems*, 77:18–31, 2005.

[19] S. Ho, E. Bullitt, and G. Gerig. Level-set evolution with region competition: automatic 3d segmentation of brain tumors. *In Proceedings of the 16th IEEE International Conference on Pattern Recognition*, 1:532–535, 2002.

[20] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, pages 321–331, 1987.

[21] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, , and A. Yezzy. Gradient flows and geometric active contour models. *in Proceedings of the 5th International Conference on Computer Vision (ICCV 95)*, pages 810–815, 1995.

[22] A. Langville, C. D. Meyer, R. Albright, J. Cox, and D. Duling. Algorithms, initializations, and convergence for the nonnegative matrix factorization. *NCSU Technical Report Math 81706*, pages 1–18, 2006.

[23] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. *Proc Conf Neural Information Processing Systems*, pages 556–562, 2001.

[24] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–793, 1999.

[25] C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore. A level set method for image segmentation in the presence of intensity inhomogeneities with application to mri. *IEEE Transactionson on Image Processing*, 20(7):2007–2016, 2011.

[26] C. Li, C. Y. Kao, J. C. Gore, and Z. Ding. Minimization of region-scalable fitting energy for image segmentation. *IEEE Transactions on Image Processing*, 17(10):1940–1949, 2008.

[27] C. Li, C. Xu, C. Gui, and M. Fox. Level set evolution without re-initialization: A new variational formulation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:430–436, 2005.

[28] C. Li, C. Xu, C. Gui, and M. D. Fox. Distance regularized level set evolution and its application to image segmentation. *IEEE Transactions on Image Processing*, 19:3242–3254, 2010.

[29] Y. Li and A. Ngom. The non-negative matrix factorization toolbox for biological data mining. *Source code for biology and medicine*, 8(1), 2013.

[30] R. Malladi, J. A. Sethian, , and B. C. Vemuri. Shape modeling with front propagation: a level set approach. *IEEE transactions on pattern analysis and machine intelligence*, 17:158–175, 1995.

[31] S. Osher and R. P. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer-Verlag, 2002.

[32] S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging, Vision, and Graphics*. Springer, 2003.

[33] S. Osher and J. Sethian. Fronts propagating with curvature dependent speed: algorithms based on hamilton-jacobi formulations. *Journal of computational physic*, 79(1):12–49, 1988.

[34] N. Passat, C. Ronse, J. Baruthio, J.-P. Armspach, C. Maillot, and C. Jahn. Region-growing segmentation of brain vessels: an atlas-based automatic approach. *Journal of Magnetic Resonance Imaging*, 21:715 – 725, 2005.

[35] D. Rivest-Henault and M. Cheriet. Unsupervised mri segmentation of brain tissues using a local linear model and level set. *Magnetic Resonance Imaging*, 29(2):243–259, 2011.

[36] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13:146 – 168, 2004.

[37] J. G. Sied, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Transactions on Medical Imaging*, 17:87–97, 1998.

[38] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.

[39] A. Tsai, A. Yezzi, and A. S.Willsky. Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation and magnification. *IEEE Transactions on Image Processing*, 10:1169–1186, 2001.

[40] U. Vovk, F. Pernus, and B. Likar. A review of methods for correction of intensity inhomogeneity in mri. *IEEE Transactions on Medical Imaging*, 26:405–21, 2007.

[41] L. Wang, L. He, A. Mishra, and C. Li. Active contours driven by local gaussian distribution fitting energy. *Signal Processing*, 89(12):2435–2447, 2009.

[42] S. K. Warfield, F. A. Kaus, M. Jolesz, and R. Kikinis. Adaptive template moderated spatially varying statistical classification. *in Proceedings of the Medical Image Computing and Computer Assisted Intervention (MICCAI 98)*, pages 431–438, 1998.

[43] T. Weglinski and A. Fabijanska. Brain tumor segmentation from mri data sets using region growing approach. *in Proceedings of the 7th International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH 11)*, pages 185 – 188, 2011.

[44] W. M. Wells, W. L. Crimson, R. Kikinis, and F. A. Jolesz. Adaptive segmentation of mri data. *IEEE Transactions on Medical Imaging*, 15:429 – 442, 1996.