Syracuse University
SURFACE

Electrical Engineering and Computer Science Technical Reports

L.C. Smith College of Engineering and Computer Science

2-4-2011

Performance Limit of Image Segmentation Algorithms

Renbin Peng
Syracuse University, pengrenbin@hotmail.com

Pramod Varshney
Syracuse University, varshney@syr.edu

Follow this and additional works at: http://surface.syr.edu/eecs_techreports

Part of the Computer Sciences Commons

Recommended Citation
Peng, Renbin and Varshney, Pramod, "Performance Limit of Image Segmentation Algorithms" (2011). *Electrical Engineering and Computer Science Technical Reports.* Paper 23.
http://surface.syr.edu/eecs_techreports/23

This Report is brought to you for free and open access by the L.C. Smith College of Engineering and Computer Science at SURFACE. It has been accepted for inclusion in Electrical Engineering and Computer Science Technical Reports by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

Department of Electrical Engineering and Computer Science

Technical Report

SYR-EECS-2011-02 Feb. 4, 2011

Performance Limit of Image Segmentation Algorithms

Renbin Peng, *Student Member, IEEE* rpeng@syr.edu
Pramod K. Varshney, *Fellow, IEEE* varshney@syr.edu

ABSTRACT: Image segmentation is a very important step in image analysis, and performance evaluation of segmentation algorithms plays a key role both in developing efficient algorithms and in selecting suitable methods for the given tasks. Although a number of publications have appeared on segmentation methodology and segmentation performance evaluation, little attention has been given to statistically bounding the performance of image segmentation algorithms. In this paper, a modified Cramér–Rao bound combined with the Affine bias model is employed to determine the performance limit of image segmentation algorithms. A fuzzy segmentation formulation is considered, of which hard segmentation is a special case. Experimental results are obtained where we compare the performance of several representative image segmentation algorithms with the derived bound on both synthetic and real-world image data.

KEYWORDS: Image segmentation, Cramér–Rao bound, Affine bias mode

Syracuse University - Department of EECS,
4-206 CST, Syracuse, NY 13244
(P) 315.443.2652 (F) 315.443.2583
http://eecs.syr.edu

Performance Limit of Image Segmentation Algorithms

Renbin Peng, *Student Member*, *IEEE*, and Pramod K. Varshney, *Fellow*, *IEEE*[1]

***Abstract*—Image segmentation is a very important step in image analysis, and performance evaluation of segmentation algorithms plays a key role both in developing efficient algorithms and in selecting suitable methods for the given tasks. Although a number of publications have appeared on segmentation methodology and segmentation performance evaluation, little attention has been given to statistically bounding the performance of image segmentation algorithms. In this paper, a modified Cramér–Rao bound combined with the Affine bias model is employed to determine the performance limit of image segmentation algorithms. A fuzzy segmentation formulation is considered, of which hard segmentation is a special case. Experimental results are obtained where we compare the performance of several representative image segmentation algorithms with the derived bound on both synthetic and real-world image data.**

***Index Terms*— Image segmentation, Cramér–Rao bound, Affine bias model**

I. Introduction

Image segmentation plays a critical role in image analysis. It subdivides an image into its constituent parts in order to extract information regarding objects of interest, and has an impact on all the subsequent image analysis tasks, such as object classification and scene interpretation [1]. Image segmentation is a challenging problem in computer vision, and a wide variety of methodologies for it have been presented, which include thresholding techniques [2], Markov random fields (MRF)-based approaches [3][4], multi-resolution algorithms [5] and partial differential equations (PDE)-based methods [6]. Surveys of image segmentation techniques can be found in [1][7]. Based on the image information being employed for the segmentation task, image segmentation algorithms can be classified into three categories: region-based segmentation, boundary- or edge-based segmentation and the methods combining both region and

[1] Renbin Peng and Pramod K. Varshney are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, 13244. (Emails: rpeng@syr.edu, varshney@syr.edu; phone: (315) 443-1060; fax: (315) 443-4745; mailing address: 4-206 Center for Science and Technology, Syracuse University, Syracuse, NY, 13244.)

boundary (edge) information.

Region-based segmentation methods aim at exploiting the image contextual information, such as spatial dependency or spatial distribution. The segmented images are expected to consist of regions within which the image content is homogeneous, while the contrast between neighboring regions is high. Typical methods falling into this category include region growing, watershed, some MRF-based methods [3], mean-shift [8] and the lossy data compression-based approach [9]. Segmentation methods based on the boundary or edge information are designed to exploit the discontinuity of the image features, such as the difference in texture or pixel intensity, on the two sides of the boundary. Typical methods in this group include gradient-based methods, such as the Canny edge detector [10], line detection methods, such as the Hough transform [11], those taking into account the interaction between boundaries (or edges) [12][13], and the methods based on physics models [14][15]. There also exist algorithms that combine region-based and boundary-based segmentations in order to benefit from fusing these two complementary approaches. There are two types of algorithms that belong to this category. The first type of algorithms carry out region and boundary segmentations sequentially [16][17], where one segmentation method is employed as the preprocessing or initialization step of another. The second type of algorithms perform segmentation by considering region and boundary information simultaneously [18][19].

While development of efficient segmentation algorithms is highly desirable, the assessment of their performance is also very important. There are basically three groups of methods for segmentation evaluation [1]. These include analysis methods, empirical goodness methods and empirical discrepancy methods [1]. The analysis methods treat the algorithms for segmentation directly, such as the evaluation of the convergence rate, the computation speed and the reasonability of the objective function design. Empirical goodness methods judge the segmented image so as to indirectly assess the performance of algorithms using quantities such as intra-region uniformity, inter-region contrast and region shape. Empirical discrepancy methods compare the segmented image with the reference image and use their difference to evaluate the performance of algorithms. For instance, position and number of mis-segmented pixels and feature values of segmented objects are all performance indicators falling into this

class. Surveys of the evaluation techniques for image segmentation can be found in [1][20][21].

Much progress has been made recently in evaluating the segmentation results, but performance of such methods tends to vary as widely as the techniques themselves. As a result, the performance of the evaluation methods is far from being satisfactory. In [1], the authors listed some of the factors which limit the advancement of evaluation methods and, in turn, the performance improvement of segmentation algorithms. These factors include a lack of common mathematical models or general strategy for evaluation, the challenges in defining wide-ranging performance metrics and statistics, the difficulties in defining the ground truth, large costs in performing comprehensive evaluations and the fact that the testing data are not representative enough for actual applications.

We note that given a specific image, among all the factors possibly affecting the performance assessment of segmentation algorithms, the most important factor is the image content. Therefore, an investigation of the performance bound, which is only associated with the available image data and is independent of the segmentation algorithms, will be very helpful to evaluate the efficiency of image segmentation techniques. A tight performance bound can tell us what the best achievable performance of any image segmentation algorithm is for the specific image content. Thus, performance bounds can serve as benchmarks for the image dataset and segmentation algorithms. They can also be used to study how the image content or image preprocessing operations affect segmentation performance. The gap between the actual segmentation error of an approach and a tight bound can provide us with the efficiency of that segmentation approach and available room for improvement.

There do exist efforts on bounding the segmentation performance from a statistical perspective. The work in [22] is based on the finite normal mixture (FNM) model assumption, where the model parameters, means and variances, are estimated using Expectation-Maximization (EM) and Classification-Maximization (CM) algorithms. Cramér–Rao bounds (CRB) on the variances of these estimates are derived. However, the FNM model is not universally applicable to all the images, and also, the unbiased estimator assumption made in [22] does not hold for many real-world segmentation algorithms, which

will be seen in our experimental results. While studying multi-spectral image segmentation [23], the performance of the Markov random fields (MRFs)-based segmentation algorithms was predicted using false alarm rate which was based on Rissanen's minimum description length (MDL) criterion. The analysis in [24] covered many detailed scenarios of segmentation, but the computational complexity, the MRF-based assumption and the use of multi-spectral image data constrained its application. In [24], the true segmentation label and two performance level parameters (sensitivity and specificity) were estimated using the EM algorithm. This scheme did not decouple the performance bound, i.e., the best achievable segmentation result for the given image data, from the specific segmentation algorithm, i.e., the EM algorithm used in [24]. In addition, the EM algorithm only guarantees to yield a locally optimal solution, which may not be appropriately used as a performance benchmark or bound, a global concept.

In this paper, we formulate image segmentation as a statistical parameter estimation problem and derive CRB on the performance measure, namely on the mean square error (MSE) of the resulting pixel labels, based on the biased estimator assumption and Affine bias model. In addition, an approximation is made when computing the expectation of the inverse Fisher information matrix to reduce the computational burden. Bootstrapping technique and empirical approximation to the second-order statistics are employed to overcome the difficulty that the probability distribution of the images is unknown. Our final goal is to derive a tight performance bound for the image segmentation problem and compare the bound with the performance of various segmentation algorithms when applied to different image datasets. The effect of the factors, such as the intensity contrast in an image on the segmentation result, are investigated via the bound, which give us insights into the achievable accuracy of a segmentation algorithm in segmenting a specific image.

This paper is structured as follows. In Section II, the image segmentation problem is shown to fit the varying coefficient model (VCM) [25] and image segmentation is formulated as a parameter estimation problem. In order to derive the biased bound later, the CRB based on the unbiased estimator assumption is discussed in Section III as a necessary intermediate step. In Section IV, the biased bound and the

optimum parameters for the Affine bias model are determined, where the methods used to calculate the bound are also discussed. In Section V, the derived biased bound is compared with several representative image segmentation algorithms using synthetic and real-world image data. We also show in Section V the comparison of these segmentation algorithms with the unbiased bound, and demonstrate the unsuitability of the unbiasedness assumption. Concluding remarks and suggestions for future work are provided in Section VI.

II. Problem Formulation

Image segmentation is a very challenging problem, and many segmentation algorithms have been proposed. However, there is a fundamental question to be asked as to whether there exists a theoretical limit to image segmentation performance and, more importantly, how much room do we have to improve the existing algorithms. In this section, as a first step to attempt to answer this question, we model the image segmentation problem as a linear estimation problem using a VCM, where the parameters of interest, i.e., the pixel labels indicating which region a pixel belongs to, are considered to be the coefficients of the VCM.

A. Varying-Coefficient Model [25]

In this subsection, we briefly introduce the VCM. Consider a random variable s whose distribution is dependent on a parameter η. In the VCM, η can be expressed as

$$\eta = F_0 + h_1 F_1(\chi_1) + \cdots + h_M F_M(\chi_M) \tag{1}$$

where $h_1, h_2, \cdots, h_M$ and $\chi_1, \chi_2, \cdots, \chi_M$ are known as the predictors for η, and $F_1, F_2, \cdots, F_M$ are functions that enable the representation of η. F_0 is the intercept term. Thus, the model is linear in the regressors, while their coefficients are allowed to change smoothly with the value of other variables which we call "effect modifiers". η is called the linear predictor, which is related to the mean $\Lambda = E\{s\}$ via the link

function $\eta = \kappa(\Lambda)$. In the simplest case of the Gaussian model, $\kappa(\Lambda) = \Lambda$ and the data s is normally distributed with mean η , and model (1) has the form

$$s = h_1 F_1(\chi_1) + \cdots + h_M F_M(\chi_M) + \varepsilon \tag{2}$$

where $E\{\varepsilon\} = 0$, $\mathrm{var}(\varepsilon) = \sigma_\varepsilon{}^2$. Other commonly used models are log-linear models, for which $\eta = \log(\Lambda)$ and s has a Poisson distribution, and the linear logistic model with $\kappa\{\Lambda\} = \log\{\Lambda/(1-\Lambda)\}$ and s is a binomial variable. A special case occurs when χ_k 's are the same variable, such as time, age or pixel coordinates as used in our work.

There are many ways to model the functions $F_k(\chi_k)$. For example, we could use flexible parametric representations, such as Fourier series, piecewise polynomials, or otherwise and more generally nonparametric functions. In our work, the *B*-spline function (tensor product *B*-splines) is employed.

B. Image Segmentation Model

In this subsection, we model the image segmentation problem using VCM. Suppose we have an image with N pixels whose observed intensity values are $y(x)$, where x are pixel indices and ordered through zig-scanning, starting from the top-left to bottom-right in an image, and $x = 1, 2, \cdots, N$. The image segmentation problem can be formulated, based on Gaussian model (2), as

$$\begin{aligned} y(x) &= s(x) + w'(x) \\ &= \left[h_1(x)F_1(x) + \cdots + h_M(x)F_M(x) + \varepsilon\right] + w'(x) \\ &= h_1(x)F_1(x) + \cdots + h_M(x)F_M(x) + w(x) \end{aligned} \tag{3}$$

where $s(x)$ are the noise-free intensity values of the pixel x. This model has the signal effect modifying variable x , where M is the number of segmented regions, and $M \le N$. (Note that the pixels which have the same features or characteristics should be classified into the same class, but these pixels classified into the same class need not be connected to each other, that is, they may be located in separate regions. The method used to calculate the bound in this paper is based on regions, not on classes, so we will consider

regions one by one, no matter whether they belong to the same class or not.) $h_k(x)$ is the pixel label of *x*, which can be considered as the membership function, representing the degree to which the pixel *x* belongs to the kth region, $0 \le h_k(x) \le 1$ and $\sum_{k=1}^{M} h_k(x) = 1$ for every x. In the rest of the paper, the terms "label value" and "membership function value" will be used interchangeably. This definition enables the model to represent a general image segmentation scenario, i.e., fuzzy segmentation [26] where each pixel can belong to different regions at the same time. As a special case of fuzzy segmentation, a pixel in hard or crisp segmentation has the membership function $h_k(x) \in \{0,1\}$. In addition to providing a more general formulation, another important reason to study fuzzy segmentation is that the CRB fails to limit the MSE if the space of a parameter becomes finite [27], i.e., the hard segmentation case.

In (3), the noise term $w(x)$ consists of two parts, the image noise $w'(x)$ and the smoothing error ε. We assume that we have a very powerful smoother and the smoothing error is very small compared with the additive noise, so the image noise dominates the noise term, i.e., $w(x) = \varepsilon + w'(x) \approx w'(x)$. In this work, the noise is considered to be independent and identically distributed (*i.i.d.*) Gaussian random variable with zero mean and variance σ^2. Also, in our work, $F_k(x)$ is modeled using the 2D *B*-spline function with the coefficient vector β_k. Let $F_k(x) = \phi(x;\beta_k)$ represent the intensity of the pixel *x* in the kth region, and $\phi(x;\beta_k) = \sum_{l=1}^{m} \beta_{kl} b_l(x)$, where $b_l(x)$ are *B*-spline basis functions and *m* is the number of knots in an image. *l* is the index of the knots which are ordered through zig-scanning starting from the top-left to bottom-right in an image. For simplicity, the knots are uniformly deployed on the entire image plane.

Thus, (3) can be written in a matrix form as

$$\begin{aligned} y(x) &= h(x)^T \cdot \phi(x;\beta) + w(x) \\ &= h(x)^T \cdot \beta \cdot b(x) + w(x) \end{aligned} \tag{4}$$

where *T* denotes the matrix transpose, $h(x) = [h_1(x), h_2(x), \cdots, h_M(x)]^T$ and $\phi(x;\beta) = [\phi(x;\beta_1), \phi(x;\beta_2), \cdots, \phi(x;\beta_k), \cdots, \phi(x;\beta_M)]^T$. Here, $\phi(x;\beta_k) = \beta_k{}^T \cdot b(x)$, where $\beta_k = [\beta_{k1}, \beta_{k2}, \cdots, \beta_{km}]^T$ and

$b(x)=[b_1(x),b_2(x),\cdots,b_m(x)]^T$. So $\phi(x;\beta)=[\beta_1{}^T,\beta_2{}^T,\cdots,\beta_M{}^T]^T\cdot b(x)=\beta\cdot b(x)$, where $\beta=[\beta_1{}^T,\beta_2{}^T,\cdots,\beta_M{}^T]^T$.

We note that a similar formulation has been used in [28][29] for developing image segmentation algorithms. In [28][29], $h_i(x)$ is considered to be equal to or very close to 0 or 1, that is, hard segmentation, while in our formulation we consider a more general segmentation configuration, i.e., fuzzy segmentation where $h_i(x)$ lies in [0,1]. In addition, in [28][29], it was argued that the pixel label, with the given Gibbsian distribution as the prior, is independent of the image content represented by β. In contrast, we do not make any assumptions on the dependence or the prior distribution.

There are several advantages to represent the image using the smoothing coefficients β, instead of the original pixel intensity information: (i) we can denote regions with various shapes and sizes, i.e., different number of pixels, using a "uniform" representation, i.e., the basis $b(x)$ and the smoothing coefficients β_k with known or controllable dimensions. Thus, the segmentation problem can be conveniently represented by some linear models, like VCM, and the analysis can be simplified; (ii) smoothing can reduce the impact of a small number of pixels with large difference in intensity from their neighboring pixels, i.e., outliers, so as to enhance the homogeneity of the image regions. It is also helpful in reducing the possibility of yielding regions with very small size, i.e., region with very few pixels; (iii) spatially varying intensity and interactions between the neighboring image areas can be taken into consideration by the smoothing representation to some extent; (iv) the smoothing procedure can represent the image content using much smaller number of coefficients compared with the number of original image pixels, and, therefore, simplifies the computation.

From (4), we can see that there are two sets of parameters $h(x)$ and β in the model, but we are only interested in the estimation of $h(x)$. We pack $h(x)$ into a large vector H and obtain $H=[h_1(1),h_2(1),\cdots,h_M(1),h_1(2),h_2(2),\cdots,h_M(2),\cdots,h_1(N),h_2(N),\cdots h_M(N)]^T$. In this paper, we assume that the segmentation algorithms are biased estimators, that is, the output, $\hat{h}(x)$, of a segmentation algorithm is a biased estimator of the true pixel label $h(x)$. More details about this

assumption as well as its justification can be found in Section IV and Appendix C. Before deriving the MSE bound under the biased estimator assumption, we first discuss the Fisher information matrix and the bound based on the unbiased estimator assumption in the next section, where the segmentation algorithm is assumed to yield an unbiased estimate of the true pixel label. We will see that the bound under the unbiasedness assumption is very useful in finding the bound under the biasedness assumption and is also helpful in the experimental part to verify the validity of the biased estimator assumption.

III. Fisher Information and Cramér–Rao Bound for Unbiased Estimator

In this section, we derive the Fisher information matrix and the Cramér–Rao bound based on the unbiased estimator assumption.

For an estimation problem with two unknown parameters, like H and β in our work, one parameter, say, H, can be considered to be the wanted parameter and the other one, β, can be considered as the unwanted one. Both of them are assumed to be random. Based on this formulation, the performance of four variations of the Bayesian bound for estimating the wanted parameter was compared in [30][31]. However, determination of all of the bounds requires either the computation of derivatives and expectation over the joint probability distributions of the observation Y and the wanted parameter or the observation and the whole parameter set, i.e., $P(Y,H)$ or $P(Y,H,\beta)$, which is a very challenging task given the variety of image contents. Here $Y=[y(1),\cdots,y(N)]^T$.

In our work, we assume H and β to be random so as to find a bound with reasonable complexity. We first determine the conditional CRB given H and β, and then find the expectation of the conditional bound with respect to H and β to obtain the global one. We will see that during the computation of expectation it is not necessary to determine the joint probability $P(H,\beta)$ and to even consider the potential dependence between H and β.

A. Fisher Information Matrix

In this subsection, we derive the Fisher information matrix conditioned on H and β, and propose a scheme to deal with the singularity of the matrix which may exist in the single image segmentation scenario. Assume that the noise $w(x)$ is *i.i.d.* Gaussian random variable with zero mean and variance σ^2, and the observed pixel intensity is also *i.i.d.* given the membership function H and the smoothing coefficient β. Then the conditional *pdf* of the observation is

$$P(Y;H,\beta)=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(\frac{-\sum_{x=1}^{N}[y(x)-h(x)^T\cdot\beta\cdot b(x)]^2}{2\sigma^2}\right) \tag{5}$$

So the log likelihood function is given by

$$L=\ln[P(Y;H,\beta)]=-\frac{N}{2}\ln 2\pi\sigma^2-\frac{1}{2\sigma^2}\sum_{x=1}^{N}[y(x)-h(x)^T\cdot\beta\cdot b(x)]^2 \tag{6}$$

We are only interested in estimating H and assume that the information about β is available, which can be estimated from the image contents and the ground-truth segmentation results. This assumption on the availability of β is helpful in simplifying the determination of the bound and also in eliminating the ambiguity in model (4) due to the multiplication of H and β. So we focus on the Fisher information matrix corresponding to H and obtain

$$[J_F(H)]_{ij}=E_{Y|H,\beta}\left\{\left[\frac{\partial L}{\partial h_i}\right]\left[\frac{\partial L}{\partial h_j}\right]^T\right\} \tag{7}$$

The detailed derivation and the resulting Fisher information matrix are provided in Appendix A.

We notice from (A.6) that $J_F(H)$ is singular, which can be verified by multiplying the first row of $J_F(H)$ by $\beta_2^T b(1)$ and the second row by $\beta_1^T b(1)$. This is because the dimension of H is usually higher than the available observation Y, especially for the case of single image segmentation, which can be seen more clearly from (4). For multi-spectral image segmentation, there may not exist such a problem, since

we have more observed image data, and the resulting Fisher information matrix for this case is shown in Appendix B. In this paper, we focus on the derivation of the bound for the segmentation of single images, and the bound for multi-spectral image segmentation can be derived in a similar manner.

To overcome the singularity problem, we transform the multi-region segmentation problem, where $M>2$, to a binary-region segmentation problem, i.e., $M=2$, by maintaining the information regarding the region of interest, say, the ith region, and by considering the remaining regions as a single "super" region. That is, the membership functions and the smoothing coefficients corresponding to the pixels in the ith region remain fixed, and the rest of the regions are merged to form a "super" region whose membership functions and the smoothing coefficients are recalculated based on the image contents of the "super" region. Thus, the segmentation model (4) can be written as

$$\begin{aligned} y(x) &= h(x)^T \cdot \beta \cdot b(x) + w(x) \\ &= h_i(x) \cdot \beta_i^T \cdot b(x) + \left[\sum_{j=1, j\neq i}^{M} \left(h_j(x) \cdot \beta_j^T \right) \right] \cdot b(x) + w(x) \\ &= h_i(x) \cdot \beta_i^T \cdot b(x) + h_{i^S}(x) \cdot \beta_{i^S}^T \cdot b(x) + w(x) \\ &= h_i(x) \cdot (\beta_i^T - \beta_{i^S}^T) \cdot b(x) + \beta_{i^S}^T \cdot b(x) + w(x) \end{aligned} \tag{8}$$

where $h_i(x)$ and β_i are the original parameters of the ith region, and $h_{i^S}(x)$ and $\beta_{i^S}^T$ correspond to the "super" region. $h_{i^S}(x) \cdot \beta_{i^S}^T \cdot b(x) = \left[\sum_{j=1, j\neq i}^{M} \left(h_j(x) \cdot \beta_j^T \right) \right] \cdot b(x)$, with $h_{i^S}(x) \geq 0$, and $h_i(x) + h_{i^S}(x) = 1$, $i = 1, 2, \cdots, M$.

Based on (8), the Fisher information matrix of $H_i = [h_i(1), \cdots, h_i(N)]$, corresponding to the ith region, can be calculated as (9), by following a similar procedure as in Appendix A but with the "super" region considered.

$$J_F(H_i) = \frac{1}{\sigma^2} \begin{bmatrix} \left(\beta_i^T b(1) - \beta_{i^s}^T b(1)\right)^2 & 0 & \cdots & 0 & 0 \\ 0 & \left(\beta_i^T b(2) - \beta_{i^s}^T b(2)\right)^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \left(\beta_i^T b(N-1) - \beta_{i^s}^T b(N-1)\right)^2 & 0 \\ 0 & 0 & \cdots & 0 & \left(\beta_i^T b(N) - \beta_{i^s}^T b(N)\right)^2 \end{bmatrix}_{N \times N} \tag{9}$$

which is not singular if $\beta_i^T b(x) - \beta_{i^s}^T b(x) \neq 0$. Since the resulting bound also requires the determination of the expectation of $\beta_i^T b(x) - \beta_{i^s}^T b(x)$ with respect to β, which will be seen in (14), we discuss the invertibility of the Fisher information matrix in the next subsection.

Thus, for $\beta_i^T b(x) - \beta_{i^s}^T b(x) \neq 0$, we have

$$J_F^{-1}(H_i) = \sigma^2 \begin{bmatrix} \frac{1}{\left(\beta_i^T b(1) - \beta_{i^s}^T b(1)\right)^2} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\left(\beta_i^T b(2) - \beta_{i^s}^T b(2)\right)^2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\left(\beta_i^T b(N-1) - \beta_{i^s}^T b(N-1)\right)^2} & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{\left(\beta_i^T b(N) - \beta_{i^s}^T b(N)\right)^2} \end{bmatrix}_{N \times N} \tag{10}$$

The same result can be obtained using the constrained CRB [32] with the "super" region scheme where the constraint is $h_i(x) + h_{i^S}(x) = 1$.

B. Cramér–Rao Bound for Unbiased Estimator

In this subsection, we derive the Cramér–Rao bound under the unbiased estimator assumption, and employ Jensen's inequality for matrix measures [33] to simplify the expectation determination procedure. We assume that the segmentation algorithms yield unbiased estimates of the pixel labels. Based on the above formulation in Section III. A, the unbiased bound of multi-region segmentation can be calculated in a region by region manner. For the ith region, we calculate the Fisher information matrix $J_F(H_i)$ and its inverse $J_F^{-1}(H_i)$ which corresponds to the conditional bound of the covariance matrix of $\hat{H}_i$. We find the expectation of $J_F^{-1}(H_i)$ with respect to H and β, and obtain the global bound for $\hat{H}_i$, which is different from the bounds discussed in [30][31] as mentioned at the beginning of Section III. Repeating the procedure for all the regions and averaging the resulting bounds, we obtain the average unbiased bound for the entire image. In this way, we decompose the estimation problem with the dimensionality equal to

MN into *M* sub-problems, each of which has the dimensionality *N*, the same size as the number of observations (the total number of pixels in an image), and therefore overcome the ambiguity due to insufficient number of observations.

Now, we study the bound on the covariance of the estimate $\hat{H}$ under the unbiasedness assumption. The conditional covariance matrix of $\hat{H}_i$, i.e., $Cov(\hat{H}_i \mid H, \beta)$, for the unbiased estimator can be written as

$$Cov(\hat{H}_i \mid H, \beta) = E_{Y|H,\beta}\{(\hat{H}_i - \hat{\mu}_{\hat{H}_i|H,\beta})(\hat{H}_i - \hat{\mu}_{\hat{H}_i|H,\beta})^T\} \geq J_F^{-1}(H_i) \tag{11}$$

where $\hat{\mu}_{\hat{H}_i|H,\beta} = E(\hat{H} \mid H, \beta)$, and the corresponding conditional bound $CRB_{Unbiased}(\hat{H}_i \mid H, \beta)$ is

$$CRB_{Unbiased}(\hat{H}_i \mid H, \beta) = \mathrm{Tr}\left[J_F^{-1}(H_i)\right] = \sigma^2 \sum_{x=1}^{N} \frac{1}{\left(\beta_i^T b(x) - \beta_{i^s}^T b(x)\right)^2} \tag{12}$$

where Tr(*U*) denotes the trace of the matrix *U*.

The global bound for H_i is determined by finding the expectation of $CRB_{Unbiased}(\hat{H}_i \mid H, \beta)$ with respect to H and β, i.e., $E_{H,\beta}\left\{CRB_{Unbiased}(\hat{H}_i \mid H, \beta)\right\}$.

The average bound for the unbiased estimator for an individual region can be found by averaging the global bounds of all the regions, that is,

$$\begin{aligned} CRB_{Unbiased-Ave}\left(\hat{H}\right) &= \frac{1}{M}\sum_{i=1}^{M} E_{H,\beta}\left\{CRB_{Unbiased}(\hat{H}_i \mid H, \beta)\right\} \\ &= \frac{1}{M}\sum_{i=1}^{M} E_{H,\beta}\left\{\mathrm{Tr}\left[J_F^{-1}(H_i)\right]\right\} \\ &= \frac{1}{M}\sum_{i=1}^{M} \mathrm{Tr}\left\{E_{\beta}\left[J_F^{-1}(H_i)\right]\right\} \end{aligned} \tag{13}$$

where the last equality holds since $\mathrm{Tr}\left[J_F^{-1}(H_i)\right]$ is not a function of *H*.

In our paper, we further average $CRB_{Unbiased-Ave}\left(\hat{H}\right)$ over all the pixels in an image and the average pixel-level bound serves as the bound on the performance of image segmentation. Since H_i's have the

same dimensions, i.e., the number of pixels included in an image, we obtain the average pixel-level bound by dividing $CRB_{Unbiased-Ave}\left(\hat{H}\right)$ with the total number of pixels, *N*, in an image, which is shown in (14).

$$\begin{aligned} CRB_{P-Unbiased-Ave}\left(\hat{H}\right) &= \frac{1}{N} CRB_{Unbiased-Ave}\left(\hat{H}\right) \\ &= \frac{1}{MN}\sum_{i=1}^{M} \mathrm{Tr}\left\{E_{\beta}\left[J_F^{-1}(H_i)\right]\right\} \end{aligned} \tag{14}$$

We notice from (10) that it is not easy to find the expectation of $J_F^{-1}(H_i)$ over β, so we employ an approximation when calculating the bound, by performing the expectation operation on $J_F(H_i)$ first and then finding its inverse, i.e., $\left(E_{\beta}\left[J_F(H_i)\right]\right)^{-1}$. According to ***Theorem 4.2*** (Jensen's inequality for matrix measures) and the ***Tracial Jensen inequalities*** in [33], we have

$$E_{\beta}\left[J_F^{-1}(H_i)\right] \geq \left(E_{\beta}\left[J_F(H_i)\right]\right)^{-1} \text{and } \mathrm{Tr}\left\{E_{\beta}\left[J_F^{-1}(H_i)\right]\right\} \geq \mathrm{Tr}\left\{\left(E_{\beta}\left[J_F(H_i)\right]\right)^{-1}\right\} \tag{15}$$

where

$$\left(E_{\beta}\left[J_F(H_i)\right]\right)^{-1} = \sigma^2 \begin{bmatrix} 1/E_{\beta}\left\{\left(\beta_i^T b(1) - \beta_{i^s}^T b(1)\right)^2\right\} & 0 & \cdots & 0 & 0 \\ 0 & 1/E_{\beta}\left\{\left(\beta_i^T b(2) - \beta_{i^s}^T b(2)\right)^2\right\} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1/E_{\beta}\left\{\left(\beta_i^T b(N-1) - \beta_{i^s}^T b(N-1)\right)^2\right\} & 0 \\ 0 & 0 & \cdots & 0 & 1/E_{\beta}\left\{\left(\beta_i^T b(N) - \beta_{i^s}^T b(N)\right)^2\right\} \end{bmatrix}_{M\times N} \tag{16}$$

Thus, a looser bound is found to ease computation, which is called the modified CRB in this paper and is indicated by the superscript *Mod*. Therefore, from (14) we have

$$CRB_{P-Unbiased-Ave}^{Mod}\left(\hat{H}\right) = \frac{1}{MN}\sum_{i=1}^{M} \mathrm{Tr}\left\{\left(E_{\beta}\left[J_F(H_i)\right]\right)^{-1}\right\} \tag{17}$$

We now discuss a special situation, where $E_{\beta}\left\{\left(\beta_i^T b(x) - \beta_{i^s}^T b(x)\right)^2\right\}$ in (16) has very small values such that its inverse is very large. In this case, the resulting average CRB value might be large. We note that the very small values of $E_{\beta}\left\{\left(\beta_i^T b(x) - \beta_{i^s}^T b(x)\right)^2\right\}$ correspond to an extreme situation where two image

regions are not distinguishable at *x*. Because $E_\beta\left\{\left(\beta_i^T b(x) - \beta_{i^s}^T b(x)\right)^2\right\}$ evaluates the average intensity difference between the two regions with the center at *x* (due to the expectation operation with respect to β), it reduces the effect when the two different regions have similar pixel intensities at *x*, by making use of the intensity information of a group of pixels. Therefore, there are very few components of $E_\beta\left\{\left(\beta_i^T b(x) - \beta_{i^s}^T b(x)\right)^2\right\}$ in (16) with very small values, given that the two image regions are reasonably separable, which has also been verified by our experiments. Thus, in our work we simply ignore the contribution of the components to the bound when they have very small values. This operation yields a reasonable tight bound. However, if we do not incorporate the expectation operation when calculating the bound, the performance of the resulting bound might be deteriorated when different regions have similar pixel intensities at *x*, which can be seen in the experimental results shown in Figs. 1 (c), 2(c) and 3(c).

From (16), we can see that $E_\beta\left\{\left(\beta_i^T b(x) - \beta_{i^s}^T b(x)\right)^2\right\}$ actually measures the square of the difference between the intensities at pixel *x* contributed by the region of interest and the "super" region. It indicates the interaction between different regions at *x*. A smaller difference means a higher similarity between the two image regions. This result corresponds to the image content which is more difficult to segment apart, and the variance of the segmentation label is larger. Here, the intensity difference evaluation is carried out by using the spline coefficients and the expectation operation, and, thus, the effect of the contribution of the neighboring pixels to the intensity at *x*, i.e., the correlation between neighboring pixels, is also taken into account. It is also interesting to notice that the separability of the two regions, which is reflected by the segmentation variance, is independent of the membership values and only related to the contrast between the intensities of the neighboring regions overlapping at a pixel. Additionally, a larger noise energy, i.e., larger σ^2, has a larger negative influence on the segmentation result, which corresponds to a higher value of the bound. We can see that the bound of (17) is consistent with these intuitive expectations.

The bound (17) has been obtained under the unbiasedness assumption but as we will see in the next section that a biased estimator is a more reasonable assumption for real-world image segmentation algorithms. Therefore, the result obtained in this section is not applicable in practice. However, it will be very useful in deriving the bound for the biased estimator case.

IV. Cramér–Rao Bound for Biased Estimator

In this section, we assume the estimator of H to be biased, and derive the bound on the MSE of the segmentation results. We continue to consider the transformed binary segmentation problem in this section.

A. Cramér–Rao Bound for Biased Estimator

From both theoretical and practical points of views, unbiased estimators do not always exist. Moreover, biased estimators often have the advantage of lower MSE over unbiased ones if they exist [34]. MSE actually includes the tradeoff between bias and covariance. In addition, unbiased estimators tend to yield very large variance, especially for some ill-posed problems, such as image segmentation. Regularization is widely used to solve ill-posed problems and the resulting estimators are often biased [35]. Many state-of-the-art image segmentation algorithms are designed under a regularization framework, in which an objective function consisting of both a fidelity term and a penalty term is optimized, resulting in biased estimators.

Following the same steps as when deriving the average bound for the unbiased estimator in the last section, we first write the expression of the conditional MSE in terms of bias and covariance, as shown in (18)

$$E\left\{\left\|\hat{H}_i - H_i\right\|^2 \mid H,\beta\right\} = \left\|g(H_i)\right\|^2 + \mathrm{Tr}\left\{Cov(\hat{H}_i \mid H,\beta)\right\} \tag{18}$$

where $g(H_i) = E\left\{\hat{H}_i\right\} - H_i$ is the bias vector of $\hat{H}_i$.

Under suitable regularity conditions on $P(Y \mid H, \beta)$, the covariance of a biased estimator of $\hat{H}$ is bounded by the CRB [30]

$$Cov(\hat{H}_i \mid H, \beta) \geq A J_F^{-1}(H_i) A^T \tag{19}$$

where

$$A = I + \frac{\partial g}{\partial H} \tag{20}$$

and I is the identity matrix.

In our work, we assume that the behavior of the bias model can be approximated by an Affine function. The Affine model has been justified and employed to study the MSE bound for estimation problems in [36]. The details of the justification of the Affine bias assumption in image segmentation can be found in Appendix C. Formally, we have

$$g(H_i) = K_i H_i + u_i \tag{21}$$

where K_i and u_i are Affine parameters for the ith region. So, following the same steps as in the last section and considering (18)-(21), we have the conditional MSE bound of a biased estimator for $\hat{H}_i$ as follows

$$E\left\{\left\|\hat{H}_i - H_i\right\|^2 \mid H, \beta\right\} \geq \left(K_i H_i + u_i\right)^T \left(K_i H_i + u_i\right) + \mathrm{Tr}\left(\left(I + K_i\right) J_F^{-1}(H_i) \left(I + K_i\right)^T\right) \tag{22}$$

Therefore, the global MSE bound for $\hat{H}_i$, i.e., $CRB_{Biased}(\hat{H}_i)$, is given by

$$E\left\{\left\|\hat{H}_i - H_i\right\|^2\right\} \geq CRB_{Biased}(\hat{H}_i) = \int \left\{\left(K_i H_i + u_i\right)^T \left(K_i H_i + u_i\right) + \mathrm{Tr}\left(\left(I + K_i\right) J_F^{-1}(H_i) \left(I + K_i\right)^T\right)\right\} P(H, \beta) dH d\beta \tag{23}$$

The average MSE bound, i.e., $CRB_{Biased-Ave}\left(\hat{H}\right)$, can be found by averaging the global bound for each region, and we, therefore, obtain

$$CRB_{Biased-Ave}\left(\hat{H}\right) = \frac{1}{M} \sum_{i=1}^{M} \int \left\{\left(K_i H_i + u_i\right)^T \left(K_i H_i + u_i\right) + \mathrm{Tr}\left(\left(I + K_i\right) J_F^{-1}(H_i) \left(I + K_i\right)^T\right)\right\} P(H, \beta) dH d\beta \tag{24}$$

B. Optimum Affine Bias Model

In this subsection, we determine the optimum $\{K_i^*, u_i^*\}$ of the Affine bias model which yield the minimum value of the bound in (23), that is,

$$\{K_i^*, u_i^*\} = \arg\min_{K_i, u_i} \left\{ \int \left[(K_i H_i + u_i)^T (K_i H_i + u_i) + \mathrm{Tr}\left((I + K_i) J_F^{-1}(H_i) (I + K_i)^T \right) \right] P(H, \beta) dH d\beta \right\} \quad (25)$$

There are basically two schemes to find the solution of the optimization problem posed in (25). The first one is to assume that K_i and u_i are functions of H or β and $\{K_i^*, u_i^*\}$ are found as the solution to the following optimization problem, as discussed in [36],

$$\{K_i^*, u_i^*\} = \arg\min_{K_i, u_i} \{ MSEB(K_i, u_i, H_i) - MSEB(0,0, H_i) \} \quad (26)$$

where $MSEB(K_i, u_i, H_i) = (K_i H_i + u_i)^T (K_i H_i + u_i) + \mathrm{Tr}\left((I + K_i) J_F^{-1}(H_i) (I + K_i)^T \right)$, and $MSEB(0,0,H_i)$ corresponds to the unbiased estimator case. As derived in [36], the resulting optimum Affine bias parameters are $K_i^* = -\mathrm{Tr}\{J_F^{-1}(H_i)\} / \left(\mathrm{Tr}\{J_F^{-1}(H_i)\} + c_i \right) I$ and $u_i^* = -\mathrm{Tr}\{J_F^{-1}(H_i)\} / \left(\mathrm{Tr}\{J_F^{-1}(H_i)\} + c_i \right) v_i$, where $\|H_i - v_i\|^2 \le c_i$ for some vector v_i and scalar c_i >0. The calculation of the bound requires the expectation of the function in (23) over $P(H, \beta)$, which is usually not tractable.

We, therefore, use the second scheme, in which we assume that M_i and u_i are not functions of H and β. As a further simplification, by using the result of (15) and also observing that $(I + K_i)\left\{ E_\beta\left[J_F^{-1}(H_i) \right] - \left(E_\beta\left[J_F(H_i) \right] \right)^{-1} \right\} (I + K_i)^T$ are positive semi-definite, we obtain a modified bound $CRB_{Biased}^{Mod}(\hat{H}_i)$ for the biased estimator, which is looser than $CRB_{Biased}(\hat{H}_i)$ shown in (23). Thus, we have

$$\begin{aligned} CRB_{Biased}(\hat{H}_i) \ge CRB_{Biased}^{Mod}(\hat{H}_i) &= \int \left\{ \mathrm{Tr}\left((I + K_i) \left(E_\beta\left[J_F(H_i) \right] \right)^{-1} (I + K_i)^T \right) \right\} P(H) dH + \int (K_i H_i + u_i)^T (K_i H_i + u_i) P(H, \beta) dH d\beta \\ &= \mathrm{Tr}\left((I + K_i) \left(E_\beta\left[J_F(H_i) \right] \right)^{-1} (I + K_i)^T \right) + \int (K_i H_i + u_i)^T (K_i H_i + u_i) P(H) dH \end{aligned} \quad (27)$$

The last equality in (27) holds because $\mathrm{Tr}\left((I + K_i) \left(E_\beta\left[J_F(H_i) \right] \right)^{-1} (I + K_i)^T \right)$ is not a function of H and

$\left(K_i H_i + u_i\right)^T \left(K_i H_i + u_i\right)$ is not a function of β.

Inspired by [37], the optimum Affine model parameters can be obtained by setting the derivative of $CRB_{Biased}^{Mod}(\hat{H}_i)$ with respect to the two parameters to zero, i.e.,

$$\frac{\partial CRB_{Biased}^{Mod}(\hat{H}_i)}{\partial K_i} = 0 \text{ and } \frac{\partial CRB_{Biased}^{Mod}(\hat{H}_i)}{\partial u_i} = 0 \tag{28}$$

Thus, we obtain the optimum parameter pair

$$K_i^* = -\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + Cov\left(H_i\right)\right\}^{-1} \tag{29}$$

and

$$u_i^* = \left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + Cov\left(H_i\right)\right\}^{-1} E_{H_i}\left(H_i\right) \tag{30}$$

Substituting K_i^* and u_i^* into $CRB_{Biased}^{Mod}(\hat{H}_i)$, we obtain the modified bound for the ith region

$$CRB_{Biased}^{Mod}(\hat{H}_i)^* = \mathrm{Tr}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} - \left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + Cov\left(H_i\right)\right\}^{-1}\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\right\} \tag{31}$$

The details of the above derivation for the parameters and the bound can be found in Appendix D.

So the average MSE bound is

$$\begin{aligned} CRB_{Biased-Ave}^{Mod}(\hat{H}) &= \frac{1}{M}\sum_{i=1}^{M} CRB_{Biased}^{Mod}(\hat{H}_i)^* \\ &= \frac{1}{M}\sum_{i=1}^{M} \mathrm{Tr}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} - \left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + Cov\left(H_i\right)\right\}^{-1}\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\right\} \end{aligned} \tag{32}$$

As before, we obtain the average pixel-level MSE bound by averaging $CRB_{Biased-Ave}^{Mod}(\hat{H})$ with respect to the total number of pixels, *N*, in an image, and we have

$$\begin{aligned} CRB_{P-Biased-Ave}^{Mod}(\hat{H}) &= \frac{1}{N} CRB_{Biased-Ave}^{Mod}(\hat{H}) \\ &= \frac{1}{MN}\sum_{i=1}^{M} \mathrm{Tr}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} - \left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + Cov\left(H_i\right)\right\}^{-1}\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\right\} \end{aligned} \tag{33}$$

We notice from (33) that the decomposition of the terms containing H and β makes the solution easily computable and no explicit expression of the joint probability $P(H,\beta)$ is required. It also avoids the study of the dependence between H and β.

C. Calculation of the MSE Bound

Computation of (33) requires the determination of $E_\beta\left[J_F(H_i)\right]$ and $Cov(H_i)$ for the ith segmented region. In this subsection, we discuss the schemes to calculate these quantities.

1) Calculation of $E_\beta\left[J_F(H_i)\right]$

We notice that calculation of $E_\beta\left[J_F(H_i)\right]$ is not straightforward even if we are able to find the distribution of β, which, of course, is also a challenging task given various image contents. So we propose to use an empirical approximation to find the expectation value.

$E_\beta\left[J_F(H_i)\right]$ is a diagonal matrix, with the diagonal elements $E_\beta\left\{\left(\beta_i^T b(x)-\beta_{i^s}^T b(x)\right)^2\right\}$. Therefore, without loss of generality, we only investigate this term.

$$\begin{aligned} E_\beta\left\{\left(\beta_i^T b(x)-\beta_{i^s}^T b(x)\right)^2\right\} &= \sum_{q=1}^{m} b_q(x)^2\left\{E_{\beta_i}\left(\beta_{iq}^2\right)+E_{\beta_{i^s}}\left(\beta_{i^s q}^2\right)\right\} - 2\sum_{q=1}^{m}\sum_{t=1}^{m} E_\beta\left(\beta_{iq}\beta_{i^s t}\right) b_q(x) b_t(x) \\ &+ 2\sum_{q=1}^{m}\sum_{t=q+1}^{m} b_q(x) b_t(x)\left\{E_{\beta_i}\left(\beta_{iq}\beta_{it}\right)+E_{\beta_{i^s}}\left(\beta_{i^s q}\beta_{i^s t}\right)\right\} \end{aligned} \tag{34}$$

Thus, we may use empirical estimation to approximate the second-order statistics in (34) and thereafter find the overall expectation of $\left(\beta_i^T b(x)-\beta_{i^s}^T b(x)\right)^2$, which avoids the step of finding the probability distribution of β. More specifically, given the noise-free image with the segmentation label H, we determine H_i and H_{i^s} as well as the pixels belonging to i^{th} and i^{sth} regions. In this way, we separate an image into two layers, one corresponding to the ith region, called the ith layer, and another corresponding to the i^{sth} region, called the i^{sth} layer. In particular, if $h_i(x)=1$, the pixel intensity at x of the ith layer is set

equal to $s(x)$, which is the pixel intensity of the original noise-free image at x; if $h_i(x)=0$, the intensity at x of the ith layer is set equal to zero. Otherwise, for a fuzzy pixel $s(x)=h_i(x)s_i^{Hard}(x)+h_{i^s}(x)s_{i^s}^{Hard}(x)$ with $h_i(x)\in(0,1)$, the pixel intensity at x of the ith layer is set equal to $s_i^{Hard}(x)$. Here $s_i^{Hard}(x)$ is the "original" hard component from the ith region, which contributes to the fuzzy pixel. This is motivated by the result of (16) in Section III. B that the segmentation error is only dependent on the intensity difference between the original regions, irrespective of the membership function values. The same procedure is carried out for the i^{sth} region. We then use the tensor B-splines to find the smoothing coefficients β_i and β_{i^s} for the two layers, respectively.

As we know that the empirical statistics will be closer to the true ones if more samples from the same distribution are used. To obtain enough valid samples of β_i and β_{i^s}, we use a "non-local" technique. That is, for β_{iq}, we search the coefficients with statistics similar to β_{iq} in the ith layer. β_{iq} together with the other similar coefficients are collected to form an ensemble, and they are considered to be various realizations of the same random variable. Then the terms including the second-order statistic of β_{iq} in (34) are calculated empirically using the collected coefficients in the ensemble. For example, $E_{\beta_i}\left(\beta_{iq}{}^2\right)$ is approximated by $\left(\sum_{d=1}^{D}\beta_{iq(d)}{}^2\right)/D$, where $\beta_{iq(d)}$ is the dth collected coefficient in the ensemble of β_{iq}, and D is the number of these coefficients, i.e., the size of the ensemble. The same procedure is carried out for β_{i^s}. The second-order statistics including β_i and β_{i^s} are calculated using the collected coefficients from both the ith and i^{sth} layers.

Ignoring the approximation error, the smoothing coefficients and the pixels are two ways to represent the same image content, so we use pixel level features to search for similar coefficients because usually the number of pixels is much larger than the coefficients and, therefore, the statistics of the pixel level are more reliable. For example, when we search for coefficients similar to β_{iq}, we divide the image into

patches centered at each knot with a fixed size, that is, the location of each smoothing coefficient is at the center of the patch. A suitable metric is employed to find the patches with a similar structure to the patch centered at β_{iq}, and the corresponding smoothing coefficients will be put in the ensemble of β_{iq}. As will be seen in Section V, we will use a metric called structural similarity (*SSIM*) index [38]. In other words, we use the similarity of the patches to represent the similarity of the smoothing coefficients.

2) Calculation of $Cov(H_i)$

The analytical solution to estimate $Cov(H_i)$ requires the knowledge of the distribution of H_i which is unknown and also not easy to find. Therefore, we use the bootstrapping technique [39]. Bootstrapping is an approach for statistical inference, and used to estimate the properties of an estimator ($Cov(H_i)$ in our work) by measuring those properties when sampling from an approximating distribution. It generates the empirical distribution of the observed data by constructing a number of resamples of the observed dataset, i.e., H_i in our work, with the same size as the observed dataset. These resamples are obtained by random sampling with replacement from the original dataset. Bootstrapping procedure is independent of the distribution, and provides an indirect method to assess the properties of the distribution which determine the sample and the parameters of interest [40]. Besides, bootstrapping is robust with respect to possibly small number of samples.

In our work, random sampling with replacement is carried out L times on H_i, and we obtain L bootstrap samples, from which the covariance matrix is calculated. This procedure is repeated R times, and the resulting R calculated covariance matrices represent an empirical bootstrap distribution of $Co\hat{v}(H_i)$ obtained from the available dataset. We accept the average of the estimated covariance matrices as the estimate of $Co\hat{v}(H_i)$. From this empirical bootstrap distribution, we can derive a bootstrap confidence interval which is also the confidence interval of the estimate of the bound and can be

considered as the variance of the bootstrap estimates.

Formally, we have

$$\hat{Cov}(H_i)^r = \frac{1}{L-1}\sum_{l=1}^{L}\left(H_i^{r^l} - \hat{\mu}_{H_i^{L^r}}\right)\left(H_i^{r^l} - \hat{\mu}_{H_i^{L^r}}\right)^T \tag{35}$$

and

$$\hat{Cov}(H_i) = \frac{1}{R}\sum_{r=1}^{R}\hat{Cov}(H_i)^r \tag{36}$$

where $H_i^{r^l}$ is the lth bootstrap sample of the same size as H_i when generating the rth covariance matrix from the empirical bootstrap distribution, and $\hat{\mu}_{H_i^{L^r}}$ is the mean vector of L bootstrap samples $H_i^{r^l}$.

Repeating the above procedure of estimating $E_\beta\left[J(H_i)\right]$ and $Cov(H_i)$ for all the M regions, and plugging these results into (33), we obtain the average pixel-level MSE bound of image segmentation for the whole image. By substituting the estimated $E_\beta\left[J(H_i)\right]$ into (17), we can also obtain the average pixel-level unbiased bound, which will be used in the next section for comparison purposes.

V. Experimental Results

In this section, we verify the efficiency of the presented MSE bound by comparing it with the segmentation results of several representative image segmentation algorithms using both synthetic and real-world image data.

A. Experiment Configuration

The two synthetic images considered here include one image with hard labels and one with hybrid labels. Hybrid here means that some pixels have hard labels and others have fuzzy labels. The real-world image is a cut of a mammogram, containing micro-calcifications, which is from the Digital Database for Screening Mammography (DDSM) [41]. The micro-calcifications are identified by the radiologists,

which are used as the ground-truth in our work.

When calculating the empirical second-order statistics, we employ the *SSIM* index [38] to find similar image patches, as mentioned before. *SSIM* measures the similarity between two images using structure information, as shown in (37)

$$SSIM\ (Y_1, Y_2) = \frac{(2\mu_{Y_1}\mu_{Y_2} + C_1)(2\sigma_{Y_1Y_2} + C_2)}{(\mu_{Y_1}^2 + \mu_{Y_2}^2 + C_1)(\sigma_{Y_1}^2 + \sigma_{Y_2}^2 + C_2)} \tag{37}$$

where μ_{Y_1}, μ_{Y_2} and σ_{Y_1}, σ_{Y_2} as well as $\sigma_{Y_1Y_2}$ denote mean intensity and contrast as well as the correlation coefficient of images Y_1 and Y_2, respectively; C_1 and C_2 are constants used to avoid instabilities for very small μ or σ. The value of $SSIM(Y_1,Y_2)$ is between 0 and 1. A higher value means more similarity between two images. In our work, Y_1 and Y_2 are two image patches under comparison.

Admittedly, the patch size, the number of similar patches found for one coefficient, the spline type and even the distance between two neighboring knots have an impact on the resulting bound. We have carried out the experiments by varying these parameters over reasonable ranges and found that the following configuration yields robust and efficient bounds. The patch size is 13 by 13 pixels, the knots are deployed every 4 pixels in both horizontal and vertical directions, and the spline function is cubic *B*-spline. There are two constraints to determine the number of patches: i) the patches with the *SSIM* index larger than 0.7 are considered as patches similar to the underlying patch; ii) the first 20 patches with the largest index values are considered as similar patches if the number of patches selected by i) exceeds 20.

As a further verification of the biased estimator assumption and Affine bias model, the unbiased bound discussed in Section III is also calculated for comparison purposes.

B. Segmentation Algorithms

The algorithms for hard image segmentation include the MRF-based algorithm [42], Otsu thresholding [43][44], dynamic clustering [45], the region-based active contour model (RACM) [46], and

the multi-scale normalized cuts-based segmentation (MNCut) [47], where RACM and MNCut are more recent and can be considered as the state-of-the-art segmentation algorithms. Those for fuzzy image segmentation include fuzzy C-means [48], fuzzy *k*-nearest neighbor (fuzzy *k*-NN) [49], and the Gath-Geva algorithm [50]. These algorithms are briefly described as follows.

MRF models have been used to represent contextual information in many pixel-based segmentation problems. A statistical method, namely the maximum *a posteriori* (MAP) approach, is often used during MRF-based image segmentation, which maximizes an objective function consisting of the *a priori* density in terms of the Gibbs distribution and the conditional probability density function of the observed image data given the distribution of the segmented region [42][51]. In this paper, we model the conditional probability density function as Gaussian and its parameters are estimated from the image data in a window centered at the pixel of interest. An adaptive window size is employed to improve its performance. The label optimization procedure is interrupted by the update of the parameters of the Gibbs field. The optimization method we used is simulated annealing.

Otsu thresholding is a classical and effective method for image segmentation. It searches for the threshold that minimizes the intra-class variance.

The dynamic clustering algorithm assumes that the different image regions obey Gaussian distributions with different means and variances. The clustering or segmentation labels and the parameters of the Gaussian model are updated in an iterative manner.

RACM [46] is based on level set evolution, which aims at overcoming the difficulties of segmentation due to the intensity inhomogeneities. The authors employ a region-based active contour model which draws upon intensity information in local regions at a controllable scale. A contour and two fitting functions that locally approximate the image intensities on the two sides of the contour are defined as the data fitting energy. A variational level set formulation incorporates the energy with a level set regularization term, and then the energy minimization is carried out for the derived curve evolution equation.

MNCut [47] uses the normalized cut graph partitioning framework of image segmentation, where a graph encoding pair-wise pixel affinity is constructed and partitioned for image segmentation. The algorithm works simultaneously across the graph scales, with an inter-scale constraint to ensure communication and consistency between the segmentations at each scale, such that both coarse and fine level details are captured.

The fuzzy C-means clustering algorithm is based on the minimization of the *C-means functional* which is used as the objective function. The minimization of the *C-means functional* is a nonlinear optimization problem that can be solved by using a variety of available methods. The most popular one is a Picard iteration through the first-order conditions for the stationary points of the *C-means functional*. The algorithm yields the weighted mean of the data items that belong to a cluster, where the weights are the membership values.

Fuzzy *k*-NN is a fuzzy version of the crisp *k*-NN algorithm, in which fuzzy sets are introduced into the algorithm. The basic step of the fuzzy *k*-NN algorithm is to assign membership of a vector as a function of the vector's distance from its *k*-nearest neighbors and those neighbors' memberships in the possible classes.

The Gath-Geva algorithm uses a distance norm based on the fuzzy maximum likelihood estimates. This distance norm involves an exponential term and thus decreases faster than the inner-product norm. The membership degrees are interpreted as the posterior probabilities of selecting the ith cluster given a data point. Gath and Geva [50] reported that the fuzzy maximum likelihood estimates clustering algorithm is able to detect clusters of varying shapes, sizes and densities.

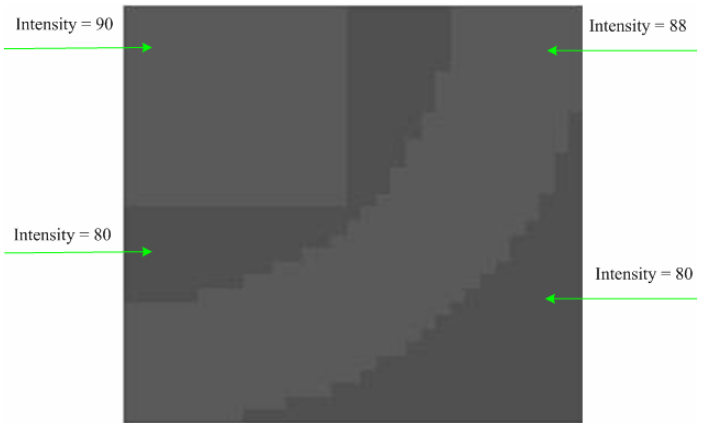
C. Experimental Results

Fig. 1 (a) shows a synthetic hard image with three intensity values, where the square in the upper-left corner has the intensity 90, the central arc has intensity 88, and the rest has intensity 80. White Gaussian

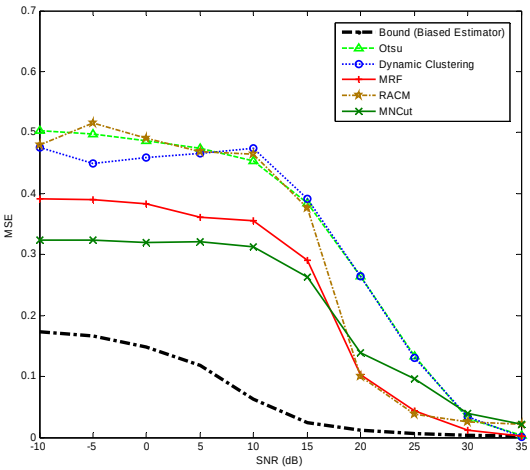
noise is added into the image with zero mean and variance σ^2. Fig. 1 (b) shows the MSE curves of the segmentation results using the above five hard image segmentation algorithms as well as the bound calculated using (33) based on the biased estimator assumption and Affine bias model. Fig. 1 (c) shows the variance curves of these segmentation algorithms and the bound calculated using (17) where we assume that the segmentation algorithms are unbiased estimators. The bounds, MSEs and variances are calculated for the particular image of Fig. 1 (a) under different noise strengths, i.e., different SNRs. At each SNR, the MSE and variance of each segmentation algorithm are the averages of 100 segmentation results. This procedure is used for all the experiments in this paper.

From Fig. 1 (b) we can see that the MSE bound (the bold dashed-dot line in the lower part of the figure) derived under the biased estimator assumption bounds the MSEs of these algorithms from below. With the increase of SNR, the bound and the MSEs decrease. When the SNR is very high, the MSEs converge to the bound. These expected results show that the bound in (33) provides a valid performance prediction of the segmentation algorithms and a benchmark of the segmentation results. In comparison, the bound in Fig. 1 (c) based on the unbiased estimator assumption, the bold dashed line, fails to bound the variance of these algorithms, which again verifies the reasonability of the biased estimator and the Affine model assumptions. In Fig. 1 (c), we use the bound values of 0.5 to represent the invalid cases where the variances calculated from the unbiased estimator assumption are very large. However, the variance should have a small value, given that the value of the pixel membership function lies in a small range of [0, 1].

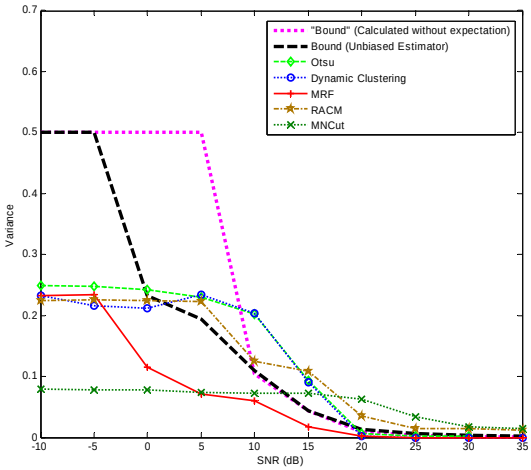
From Fig. 1 (b), we can see that the MRF-based segmentation algorithm exploits the correlation between neighboring pixels and yields a better result, in terms of smaller MSE, than the methods which consider pixels to be independent when carrying out segmentation, such as dynamic clustering. This also shows the reasonability of our representation of the image using smoothing coefficients and the expectation operation with respect to β when calculating the bound, which take into account the correlation information contained in an image. As a further verification, in Fig. 1 (c) we draw the “bound”



(a)



(b)

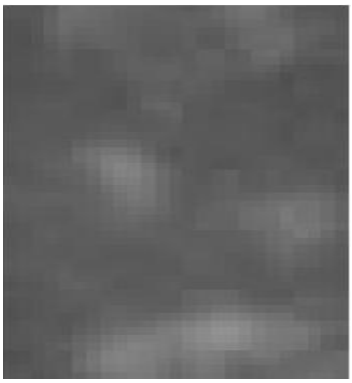


(c)

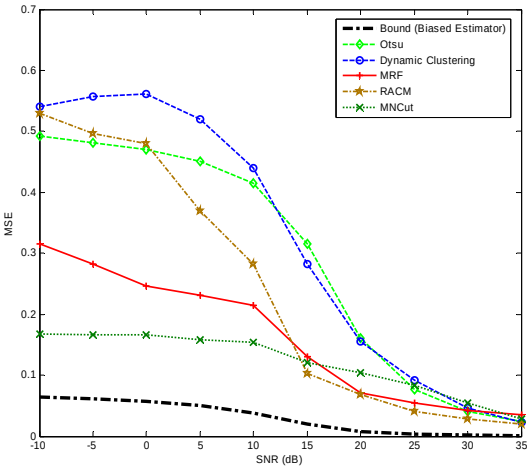
Fig. 1 Bounds for hard image segmentation (synthetic image). (a) Synthetic hard image; (b) MSEs and bound under the biased estimator assumption; (c) variances and bound under the unbiased estimator assumption.

curve, the dotted line at the right hand side of the unbiased bound, which is based on the unbiased estimator assumption but calculated by using $\left(\beta_i^T b(x) - \beta_{i^s}^T b(x)\right)^2$ directly from the pixel intensity and without the expectation operation with respect to β. We can see that not taking correlation into account yields an even worse result. Similar results can also be seen in Fig. 2 (c) and Fig. 3 (c).

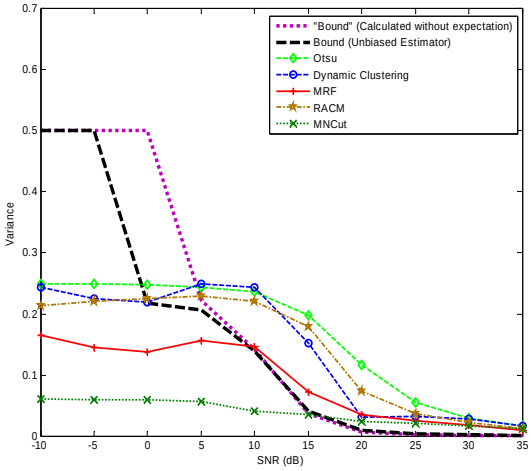
Fig. 2 shows the results when calculating the bounds and MSEs using the real-world mammogram data. We can see that the presented biased estimator-based bound performs satisfactorily in predicting the performance limit of the algorithms, while the one based on the unbiased assumption fails.



(a)



(b)



(c)

Fig. 2 Bounds for hard image segmentation (real-world image). (a) Mammogram with micro-calcifications; (b) MSEs and bound for biased estimator assumption; (c) variances and bound for unbiased estimator assumption.

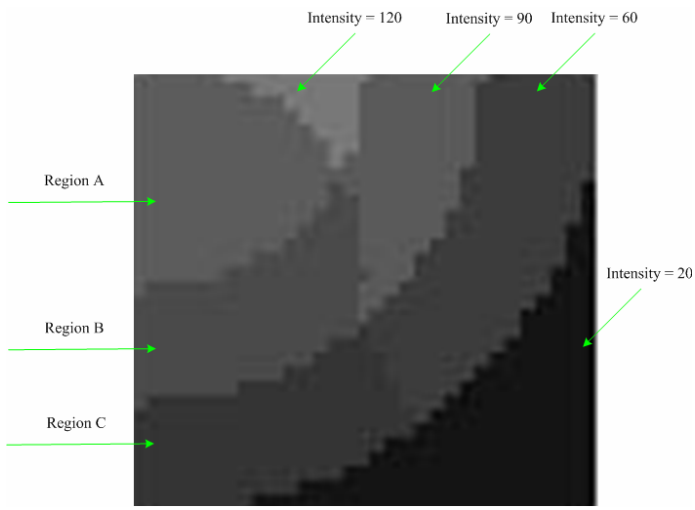
Fig. 3 deals with hybrid image segmentation for the synthetic image shown in Fig. 3 (a). There are four basic image regions, corresponding to the intensity values of 120, 90, 60 and 20, respectively. The

three arc regions at the left side of the diagonal curves are fuzzy regions, denoted as Regions A, B, and C, and have membership values of [0.5, 0.2, 0.2, 0.1], [0, 0.6, 0.3, 0.1] and [0, 0, 0.8, 0.2], respectively. The rest of the four regions are hard ones with the intensity values mentioned above. Once again white Gaussian noise is added into the image with zero mean and variance σ^2. Fig. 3 (b) shows the MSE curves of the segmentation results using the three fuzzy image segmentation algorithms when the biased estimator assumption and Affine bias model are employed. Fig. 3 (c) shows the variance curves of the segmentation algorithms, the bound calculated using (17) for the unbiased estimator assumption and the "bound" determined by ignoring the expectation operation. We can see from the figures that the bound based on the biased estimator assumption is valid but those based on the unbiased estimator assumption fail again.

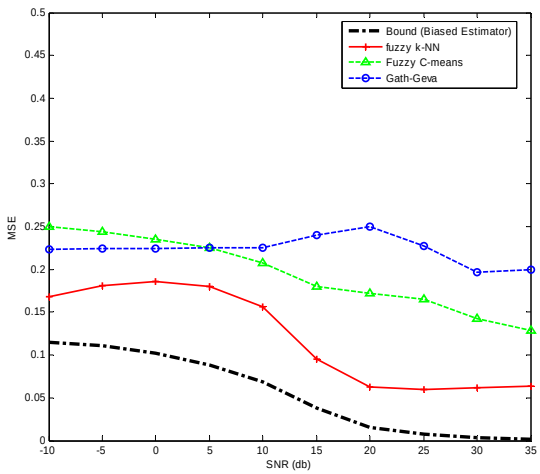
VI. CONCLUSION

Image segmentation is a very important but challenging problem for computer vision and image analysis. However, performance limits of segmentation algorithms are seldom studied from a statistical perspective. This paper developed a systematic method to estimate a lower bound on the MSE of segmentation algorithms under a statistical estimation framework. The bound was based on the biased estimator assumption and Affine bias model, where an approximation was employed to simplify the computation when determining the expectation on the inverse of the Fisher information matrix. Additionally, non-local searching and boostrapping techniques were used to approximate the unknown second-order statistics during the computation of the bound. The theoretical analysis and experimental results show that the presented bound is efficient and robust in bounding the performance of the segmentation algorithms and providing a benchmark for the segmentation problem.

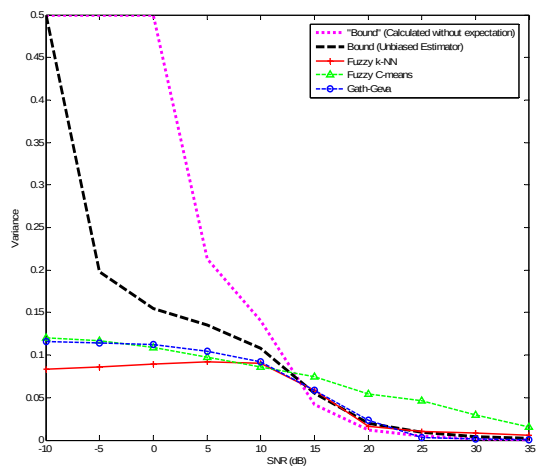
There are many future research directions that are worth pursuing. An investigation on the probability distribution estimation techniques may be helpful to improve the computation of the expectation involved



(a)



(b)



(c)

Fig. 3 Bounds for hybrid image segmentation (synthetic image). (a) Synthetic hybrid image; (b) MSEs and bound for biased estimator assumption; (c) variances and bound for unbiased estimator assumption.

in the bound, where statistical learning methods may be helpful. In our current work, we mainly discussed the problem of segmenting a single image, and only mentioned multi-spectral image segmentation in Appendix B and did not consider the 3D scenario. Future research on the extension of the developed bound to the multi-spectral and 3D images will be an interesting research topic. When developing the

bound, the ground truth information about the noise-free image and the membership value of each pixel label is required. Research on approaches which can reduce the dependence of the bound on such information will be both theoretically and practically useful. Perhaps image denoising and linear regression techniques will be helpful in handling it. Finally, the presented bound may also be useful in color images, which could be an excellent extension of our work to more real-world applications

ACKNOWLEDGEMENT

This work was supported by AFOSR under grant FA9550-06-C-0036.

APPENDIX

APPENDIX A

CALCULATING FISHER INFORMATION MATRIX (FOR SINGLE IMAGE)

Assume that the noise $w(x)$ is *i.i.d.* Gaussian random variable with zero mean and variance σ^2, and the observed pixel intensity is also *i.i.d.* given the membership H and the coefficient β. Then the conditional *pdf* of the observation is

$$P(Y;H,\beta)=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(\frac{-\sum_{x=1}^{N}[y(x)-h(x)^T\cdot\beta\cdot b(x)]^2}{2\sigma^2}\right) \tag{A.1}$$

So

$$
\begin{aligned}
L &= \ln[P(Y;H,\beta)] \\
&= -\frac{N}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{x=1}^{N}[y(x)-h(x)^T\cdot\beta\cdot b(x)]^2 \\
&= -\frac{N}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{x=1}^{N}\left\{\begin{array}{l} y(x)-[h_1(x)(\beta_{11}b_1(x)+\beta_{12}b_2(x)+\cdots+\beta_{1m}b_m(x))+h_2(x)(\beta_{21}b_1(x)+\beta_{22}b_2(x)+\cdots+\beta_{2m}b_m(x))) \\ +\cdots+h_M(x)(\beta_{M1}b_1(x)+\beta_{M2}b_2(x)+\cdots+\beta_{Mm}b_m(x))] \end{array}\right\}^2
\end{aligned}
\tag{A.2}
$$

Therefore,

$$
\frac{\partial L}{\partial h_k(x)} = \frac{1}{\sigma^2}\beta_k{}^T\cdot b(x)\cdot w(x) \tag{A.3}
$$

where $k = 1,2,\cdots,M$.

$$
\begin{aligned}
E\left\{\left[\frac{\partial L}{\partial h_k(x)}\right]\left[\frac{\partial L}{\partial h_{k'}(x')}\right]^T\right\} &= E\left\{\frac{1}{\sigma^4}\beta_k{}^T\cdot b(x)\cdot w(x)\cdot w(x')\cdot b(x')^T\cdot\beta_{k'}\right\} \\
&= \begin{cases} \frac{1}{\sigma^2}\beta_k{}^T\cdot b(x)\cdot b(x)^T\cdot\beta_{k'}, & if \quad x = x' \\ 0, & if \quad x \neq x' \end{cases}
\end{aligned}
\tag{A.4}
$$

Fisher information matrix is determined as follows,

$$
[J_F(H)]_{ij} = E_{Y|H,\beta}\left\{\left[\frac{\partial L}{\partial h_i}\right]\left[\frac{\partial L}{\partial h_j}\right]^T\right\} \tag{A.5}
$$

and

$$
J_F(H) = E_{Y|H,\beta}\left\{\left[\frac{\partial L}{\partial h_k(x)}\right]\left[\frac{\partial L}{\partial h_{k'}(x')}\right]^T\right\}_{MN\times MN} =
$$

$$
\frac{1}{\sigma^2}\left[\begin{array}{ccc|ccc|c|ccc}
\beta_1^T b(1)b(1)^T\beta_1 & \cdots & \beta_1^T b(1)b(1)^T\beta_M & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\beta_2^T b(1)b(1)^T\beta_1 & \cdots & \beta_2^T b(1)b(1)^T\beta_M & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
 & \vdots & & & \vdots & & \vdots & & \vdots & \\
\beta_M{}^T b(1)b(1)^T\beta_1 & \cdots & \beta_M{}^T b(1)b(1)^T\beta_M & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\hline
0 & \cdots & 0 & \beta_1^T b(2)b(2)^T\beta_1 & \cdots & \beta_1^T b(2)b(2)^T\beta_M & \cdots & 0 & \cdots & 0 \\
0 & \cdots & 0 & \beta_2^T b(2)b(2)^T\beta & \cdots & \beta_2^T b(2)b(2)^T\beta_M & \cdots & 0 & \cdots & 0 \\
 & \vdots & & & \vdots & & \cdots & & \vdots & \\
0 & \cdots & 0 & \beta_M{}^T b(2)b(2)^T\beta_1 & \cdots & \beta_M{}^T b(2)b(2)^T\beta_M & \cdots & 0 & \cdots & 0 \\
\hline
 & \vdots & & & \vdots & & \cdots & & \vdots & \\
\hline
0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \beta_1^T b(N)b(N)^T\beta_1 & \cdots & \beta_1^T b(N)b(N)^T\beta_M \\
0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \beta_2^T b(N)b(N)^T\beta_1 & \cdots & \beta_2^T b(N)b(N)^T\beta_M \\
 & \vdots & & & \vdots & & \cdots & & \vdots & \\
0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \beta_M{}^T b(N)b(N)^T\beta & \cdots & \beta_M{}^T b(N)b(N)^T\beta_M
\end{array}\right]
\tag{A.6}
$$

APPENDIX B

CALCULATING FISHER INFORMATION MATRIX (FOR MULTI-SPECTRAL IMAGES)

For a multi-spectral image set including P images, H is the same for all of them, b can be different if the smoothing configuration, such as the number, position and size of the spacing of knots, are different from one image to another, but β usually are different for different images. Therefore, we have the segmentation model

$$y^i(x) = h(x)^T \phi^i(x;\beta) + w^i(x) = h(x)^T \cdot \beta^i \cdot b^i(x) + w^i(x) \tag{B.1}$$

where $i = 1,2,\cdots,P$, $x = 1,2,\cdots,N$, $b^i(x) = [b_1^i(x), b_2^i(x), \cdots, b_{m_i}^i(x)]^T$, $\beta^i = [\beta^{i\,T}_1, \beta^{i\,T}_2, \cdots, \beta^{i\,T}_M]^T$, and $\beta^i_k = [\beta^i_{k_1}, \beta^i_{k_2}, \cdots, \beta^i_{km_i}]^T$. The noise may be different, so we assume $w^i(x)$ are *i.i.d.* Gaussian noise with zero mean and variance σ_i^2 . For simplicity, we use the same knot configuration for every image. Then the model is simplified to

$$y^i(x) = h(x)^T \cdot \beta^i \cdot b(x) + w^i(x) \tag{B.2}$$

We still assume that the observed pixel intensities are *i.i.d.* given the membership H and the coefficient β, so the conditional *pdf* of the observation is

$$\begin{aligned} P(Y;H,\beta^1,\beta^2,\cdots,\beta^P) &= \left(\prod_{i=1}^{P} \frac{1}{\sqrt{2\pi\sigma_i^2}}\right)^N \exp\left(-\sum_{x=1}^{N}\sum_{i=1}^{P} \frac{[y^i(x) - h(x)^T \cdot \beta^i \cdot b(x)]^2}{2\sigma_i^2}\right) \\ &= \prod_{i=1}^{P}\left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right)^N \exp\left(-\sum_{i=1}^{P} \frac{\sum_{x=1}^{N}[y^i(x) - h(x)^T \cdot \beta^i \cdot b(x)]^2}{2\sigma_i^2}\right) \end{aligned} \tag{B.3}$$

So

$$\begin{aligned} L &= \ln[P(Y;H,\beta^1,\beta^2,\cdots,\beta^P)] \\ &= -\frac{N}{2}\sum_{i=1}^{P}\ln 2\pi\sigma_i^2 - \sum_{i=1}^{P}\frac{1}{2\sigma_i^2}\sum_{x=1}^{N}[y^i(x) - h(x)^T \cdot \beta^i \cdot b(x)]^2 \\ &= -\sum_{i=1}^{P}\frac{1}{2\sigma_i^2}\sum_{x=1}^{N}\{y^i(x) - [h_1(x)(\beta_{11}^i b_1(x) + \beta_{12}^i b_2(x) + \cdots + \beta_{1m}^i b_m(x)) + \cdots + h_M(x)(\beta_{M1}^i b_1(x) + \beta_{M2}^i b_2(x) + \cdots + \beta_{Mm}^i b_m(x))]\}^2 \\ &\quad -\frac{N}{2}\sum_{i=1}^{P}\ln 2\pi\sigma_i^2 \end{aligned} \tag{B.4}$$

Therefore,

$$\frac{\partial L}{\partial h_k(x)} = \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_k{}^T \cdot b(x) \cdot w^i(x) \tag{B.5}$$

$$\begin{aligned}
E\left\{\left[\frac{\partial L}{\partial h_k(x)}\right]\left[\frac{\partial L}{\partial h_{k'}(x')}\right]^T\right\} &= E\left\{\left[\sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_k{}^T \cdot b(x) \cdot w^i(x)\right]\left[\sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_{k'}{}^T \cdot b(x') \cdot w^i(x')\right]^T\right\} \\
&= E\left\{\sum_{j=1}^{P}\sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_k{}^T \cdot b(x) \cdot w^i(x) \frac{1}{\sigma_j^2} b(x')^T \cdot \beta^j{}_{k'} \cdot w^j(x')\right\} \\
&= E\left\{\sum_{j=1}^{P}\sum_{i=1}^{P} \frac{1}{\sigma_i^2 \sigma_j^2} \beta^i{}_k{}^T \cdot b(x) \cdot w^i(x) \cdot w^j(x') \cdot b(x')^T \cdot \beta^j{}_{k'}\right\} \\
&= \begin{cases} \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_k{}^T \cdot b(x) \cdot b(x)^T \cdot \beta^i{}_{k'}, if\ \ x = x'\ and\ i = j \\ 0, if\ \ x = x' \end{cases}
\end{aligned} \tag{B.6}$$

So we have

$$J_F(H) = E_{Y|H,\beta}\left\{\left[\frac{\partial L}{\partial h_k(x)}\right]\left[\frac{\partial L}{\partial h_{k'}(x')}\right]^T\right\}_{MN \times MN} =$$

$$\begin{bmatrix}
\sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_1{}^T \cdot b(1) \cdot b(1)^T \cdot \beta^i{}_1 & \cdots & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_1{}^T \cdot b(1) \cdot b(1)^T \cdot \beta^i{}_M & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_2{}^T \cdot b(1) \cdot b(1)^T \cdot \beta^i{}_1 & \cdots & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_2{}^T \cdot b(1) \cdot b(1)^T \cdot \beta^i{}_M & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
 & \vdots & & & \vdots & & \vdots & & \vdots & \\
\sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_M{}^T \cdot b(1) \cdot b(1)^T \cdot \beta^i{}_1 & \cdots & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_M{}^T \cdot b(1) \cdot b(1)^T \cdot \beta^i{}_M & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
0 & \cdots & 0 & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_1{}^T \cdot b(2) \cdot b(2)^T \cdot \beta^i{}_1 & \cdots & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_1{}^T \cdot b(2) \cdot b(2)^T \cdot \beta^i{}_M & \cdots & 0 & \cdots & 0 \\
0 & \cdots & 0 & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_2{}^T \cdot b(2) \cdot b(2)^T \cdot \beta^i{}_1 & \cdots & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_2{}^T \cdot b(2) \cdot b(2)^T \cdot \beta^i{}_M & \cdots & 0 & \cdots & 0 \\
 & \vdots & & & \vdots & & \cdots & & \vdots & \\
0 & \cdots & 0 & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_M{}^T \cdot b(2) \cdot b(2)^T \cdot \beta^i{}_1 & \cdots & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_M{}^T \cdot b(2) \cdot b(2)^T \cdot \beta^i{}_M & \cdots & 0 & \cdots & 0 \\
 & \vdots & & & \vdots & & \cdots & & \vdots & \\
0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_1{}^T \cdot b(N) \cdot b(N)^T \cdot \beta^i{}_1 & \cdots & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_1{}^T \cdot b(N) \cdot b(N)^T \cdot \beta^i{}_M \\
0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_2{}^T \cdot b(N) \cdot b(N)^T \cdot \beta^i{}_1 & \cdots & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_2{}^T \cdot b(N) \cdot b(N)^T \cdot \beta^i{}_M \\
 & \vdots & & & \vdots & & \cdots & & \vdots & \\
0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_M{}^T \cdot b(N) \cdot b(N)^T \cdot \beta^i{}_1 & \cdots & \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \beta^i{}_M{}^T \cdot b(N) \cdot b(N)^T \cdot \beta^i{}_M
\end{bmatrix} \tag{B.7}$$

APPENDIX C

JUSTIFICATION OF THE BIASED ESTIMATOR ASSUMPTION AND AFFINE BIAS MODEL

The estimation problem in linear models was analyzed in [52]-[54]. The linear model is

$$Y = Q\theta + n \tag{C.1}$$

where Y is the observation, θ is a parameter vector, Q is a model matrix, and n is zero-mean random vector. The estimator of θ is assumed to be linear, i.e., $\hat{\theta} = GY$, which estimates θ by performing a weighted average operation over the observation. Linear estimators are quite frequently used for least squire estimation problems, whose forms have been established by solving optimization problems, with the constraints put on Q, θ and even n. These constraints can be considered as the prior information on these parameters and the penalties under the regularization framework.

Similarly, image segmentation can also be modeled as a linear estimation problem, as shown in (4)

$$\begin{aligned} y(x) &= h(x)^T \cdot \phi(x;\beta) + w(x) \\ &= h(x)^T \cdot \beta \cdot b(x) + w(x) \end{aligned} \tag{4}$$

where $\phi(x;\beta)^T$ can be considered as the model matrix and $h(x)$ is the label parameter vector to be estimated. During the segmentation procedure, some prior information about $\phi(x;\beta)$, $h(x)$ and $w(x)$ is usually employed as the penalty terms of the objective functions for segmentation, to reduce the solution space under regularization framework. For example, the smoothness assumption is often made on the labels of the neighboring pixels, like that used in the MRF-based algorithms, which equivalently brings the constraint on $h(x)$. Moreover, local information is often used during the estimation procedure, that is, $h(x)$ is often estimated by using the observation Y around the coordinate x. Thus, it is reasonable to assume that many image segmentation algorithms, especially the state-of-the-art ones, perform the label estimation using linear estimators $\hat{H} = GY$.

Here, we consider the penalty or prior information resulting from the label smoothness assumption, and assume that H forms Gaussian MRF

$$\begin{aligned} h(x_j) &= \sum_{x_{j_l} \in \eta_{x_j}} a_{j_l} h\left(x_{j_l}\right) + \omega(x_j) \\ &= \sum_{x_{j_l} \in \eta_{x_j}, x_{j_l} \neq x} a_{j_l} h\left(x_{j_l}\right) + \omega(x_j) + a_x h(x) \end{aligned} \tag{C.2}$$

where x_{j_l} denotes the indices of the lth neighbor of the pixel x_j in the neighborhood system η_{x_j} of x_j, and $\omega(x_j)$ is zero mean Gaussian noise vector. Pixel x also belongs to η_{x_j}. a_{j_l} and a_x are the model parameters. In this paper, two pixels are called neighbors if they are close to each other spatially and their observations have an impact on the estimation of the pixel labels of each other. So it is not compulsory for two neighboring pixels to be deployed in a way that one is followed immediately by another spatially.

With the neighboring information incorporated in the segmentation procedure, the linear estimator finds the weighted average over the observation in a local window. We can also consider that the weighted average is carried out over the whole set of observations in an image, but the weights decrease with the increase of the distance between the coordinates of the observations and the pixel of interest. Here, we only consider the observations which are neighbors of the pixel of interest. We have

$$\hat{h}(x) = G_x Y_x \tag{C.3}$$

where G_x and Y_x are the weighting matrix and observation vector corresponding to a neighborhood system of the pixel at *x*. More specifically, $G_x = [g_1, g_2, \cdots, g_M]^T$ and $g_i = [g_{i1}, g_{i2}, \cdots, g_{iC(x)}]^T$, where $i = 1,2,\cdots,M$, and $C(x)$ is the total number of neighboring pixels of pixel *x*. $C(x)$ is equal to the size of η_x, and may be different from pixel to pixel . $Y_x = [y(x_1), y(x_2), \cdots, y(x_{C(x)})]^T$, which is the vector consisting of the neighboring pixels of *x*.

We claim that if pixel x_j is the neighbor of pixel *x*, then pixel *x* is the neighbor of pixel x_j. Thus, substituting (4) and (C.2) into (C.3), we have

$$
\begin{aligned}
\hat{h}(x) &= G_x Y_x \\
&= \sum_j g_j y(x_j) \\
&= \sum_j g_j \left(\phi(x_j;\beta)^T \cdot h(x_j) + w(x_j)\right) \\
&= \sum_j g_j \left\{\phi(x_j;\beta)^T \cdot \left(\sum_{x_{j_l}\in\eta_{x_j}, x_{j_l}\neq x} a_{j_l} h\left(x_{j_l}\right) + \omega(x_j) + a_x h(x)\right) + w(x_j)\right\} \\
&= \sum_j g_j \left\{\phi(x_j;\beta)^T \cdot a_x h(x) + \phi(x_j;\beta)^T \cdot \left(\sum_{x_{j_l}\in\eta_{x_j}, x_{j_l}\neq x} a_{j_l} h\left(x_{j_l}\right) + \omega(x_j)\right) + w(x_j)\right\} \\
&= \sum_j g_j \left[\phi(x_j;\beta)^T \cdot a_x\right] h(x) + \sum_j g_j \left[\phi(x_j;\beta)^T \cdot \left(\sum_{x_{j_l}\in\eta_{x_j}, x_{j_l}\neq x} a_{j_l} h\left(x_{j_l}\right) + \omega(x_j)\right) + w(x_j)\right]
\end{aligned}
\tag{C.4}
$$

where g_j is the jth column of the matrix G_x and $j = 1,2,\cdots,C(x)$. $g_j\left[\phi(x_j;\beta)^T \cdot a_x\right]$ is a $M \times M$ matrix, and $\phi(x_j;\beta)^T \cdot \left(\sum_{x_{j_l}\in\eta_{x_j}, x_{j_l}\neq x} a_{j_l} h\left(x_{j_l}\right) + \omega(x_j)\right)$ is a scalar. The expected value of this linear estimator, given the true value of $h(x)$, is

$$
\begin{aligned}
E\left\{\hat{h}(x) \mid h(x)\right\} &= E\left\{\sum_j g_j \left[\phi(x_j;\beta)^T \cdot a_x\right] h(x) + \sum_j g_j \left[\phi(x_j;\beta)^T \cdot \left(\sum_{x_{j_l}\in\eta_{x_j}, x_{j_l}\neq x} a_{j_l} h\left(x_{j_l}\right) + \omega(x_j)\right) + w(x_j)\right]\right\} \\
&= E\left\{\sum_j g_j \left[\phi(x_j;\beta)^T \cdot a_x\right] h(x)\right\} + E\left\{\sum_j g_j \left[\phi(x_j;\beta)^T \cdot \left(\sum_{x_{j_l}\in\eta_{x_j}, x_{j_l}\neq x} a_{j_l} h\left(x_{j_l}\right) + \omega(x_j)\right) + w(x_j)\right]\right\} \\
&= E\left\{\sum_j g_j \left[\phi(x_j;\beta)^T \cdot a_x\right]\right\} \cdot h(x) + E\left\{\sum_j g_j \left[\phi(x_j;\beta)^T \cdot \left(\sum_{x_{j_l}\in\eta_{x_j}, x_{j_l}\neq x} a_{j_l} h\left(x_{j_l}\right) + \omega(x_j)\right) + w(x_j)\right]\right\}
\end{aligned}
\tag{C.5}
$$

So the bias vector of the linear estimator is

$$
\begin{aligned}
g\left(h(x)\right) &= E\left\{\hat{h}(x) \mid h(x)\right\} - h(x) \\
&= E\left\{\sum_j g_j \left[\phi(x_j;\beta)^T \cdot a_x\right]\right\} \cdot h(x) + E\left\{\sum_j g_j \left[\phi(x_j;\beta)^T \cdot \left(\sum_{x_{j_l}\in\eta_{x_j}, x_{j_l}\neq x} a_{j_l} h\left(x_{j_l}\right) + \omega(x_j)\right) + w(x_j)\right]\right\} - h(x) \\
&= \left(E\left\{\sum_j g_j \left[\phi(x_j;\beta)^T \cdot a_x\right]\right\} - I\right) \cdot h(x) + E\left\{\sum_j g_j \left[\phi(x_j;\beta)^T \cdot \left(\sum_{x_{j_l}\in\eta_{x_j}, x_{j_l}\neq x} a_{j_l} h\left(x_{j_l}\right) + \omega(x_j)\right) + w(x_j)\right]\right\} \\
&= K_x \cdot h(x) + u_x
\end{aligned}
\tag{C.6}
$$

where $K_x = E\left\{\sum_j g_j\left[\phi(x_j;\beta)^T \cdot a_x\right]\right\} - I$ and $u_x = E\left\{\sum_j g_j\left[\phi(x_j;\beta)^T \cdot \left(\sum_{x_{jl}\in\eta_{x_j}, x_{jl}\neq x} a_{jl} h\left(x_{jl}\right) + \omega(x_j)\right) + w(x_j)\right]\right\}$.

The subscript "*x*" of K_x and u_x means that these two quantities are relevant to pixel *x*. K_x and u_x can be further decomposed for each region type. That is, $K_x = [K_{x:1}, K_{x:2}, \cdots, K_{x:M}]^T$ and $u_x = [u_{x:1}, u_{x:2,}, \cdots, u_{x:M}]^T$. Here, for the ith region type, $K_{x:i} = [K_{x:i,1}, K_{x:i,2}, \cdots, K_{x:i,M}]^T$, a $M \times 1$ vector, and $u_{x:i}$ is a scalar, $i = 1,2,\cdots,M$.

In the "super" region scheme employed in our work, we have two regions, i.e., i^{th} and i^{sth} regions, when we consider the segmentation performance for the ith region. So, *M*=2, $K_x = [K_{x:i}, K_{x:i^s}]^T$, $u_x = [u_{x:i}, u_{x:i^s}]^T$ and $K_{x:i} = [K_{x:i,1}, K_{x:i,2}]^T$. From (C. 6) we have

$$\begin{aligned} g\left(h_i(x)\right) &= E\left\{\hat{h}_i(x) \mid h(x)\right\} - h_i(x) \\ &= E\left\{\hat{h}_i(x) \mid h_i(x), h_{i^s}(x)\right\} - h_i(x) \\ &= \left(K_{x:i,1} \cdot h_i(x) + K_{x:i,2} \cdot h_{i^s}(x)\right) + u_{x:i} \\ &= \left(K_{x:i,1} - K_{x:i,2}\right) \cdot h_i(x) + K_{x:i,2} + u_{x:i} \\ &= K'_{x:i} \cdot h_i(x) + u'_{x:i} \end{aligned} \tag{C.7}$$

where $K'_{x:i} = K_{x:i,1} - K_{x:i,2}$ and $u'_{x:i} = K_{x:i,2} + u_{x:i}$ and we have employed the relation of $h_i(x) + h_{i^s}(x) = 1$ in the derivation. Therefore, we have

$$
\begin{aligned}
g(H_i) &= E\{\hat{H}_i \mid H_i\} - H_i \\
&= E\{\hat{H}_i \mid H_i, H_{i^S}\} - H_i \\
&= \left(\begin{bmatrix} K_{1:i,1} h_i(1) \\ K_{2:i,1} h_i(2) \\ \vdots \\ K_{x:i,1} h_i(x) \\ \vdots \\ K_{N:i,1} h_i(N) \end{bmatrix} + \begin{bmatrix} K_{1:i,2} h_{i^S}(1) \\ K_{2:i,2} h_{i^S}(2) \\ \vdots \\ K_{x:i,2} h_{i^S}(x) \\ \vdots \\ K_{N:i,2} h_{i^S}(N) \end{bmatrix} \right) + \begin{bmatrix} u_{1:i} \\ u_{2:i} \\ \vdots \\ u_{x:i} \\ \vdots \\ u_{N:i} \end{bmatrix} \\
&= \begin{bmatrix} (K_{1:i,1} - K_{1:i,2}) h_i(1) \\ (K_{2:i,1} - K_{2:i,2}) h_i(2) \\ \vdots \\ (K_{x:i,1} - K_{x:i,2}) h_i(x) \\ \vdots \\ (K_{N:i,1} - K_{N:i,2}) h_i(N) \end{bmatrix} + \begin{bmatrix} K_{1:i,2} + u_{1:i} \\ K_{2:i,2} + u_{2:i} \\ \vdots \\ K_{x:i,2} + u_{x:i} \\ \vdots \\ K_{N:i,2} + u_{N:i} \end{bmatrix} \\
&= K_i \cdot H_i + u_i
\end{aligned}
\tag{C.8}
$$

where $K_i = \begin{bmatrix} (K_{1:i,1} - K_{1:i,2}) & 0 & \cdots & 0 & \cdots & 0 \\ 0 & (K_{2:i,1} - K_{2:i,2}) & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & (K_{x:i,1} - K_{x:i,2}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & (K_{N:i,1} - K_{N:i,2}) \end{bmatrix}_{N \times N}$, a $N \times N$ diagonal matrix, and $u_i = [K_{1:i,2} + u_{1:i}, \quad K_{2:i,2} + u_{2:i}, \quad \cdots, \quad K_{x:i,2} + u_{x:i}, \quad \cdots, \quad K_{N:i,2} + u_{N:i}]^T$, a $N \times 1$ vector .

From the above analysis, we can see that in many segmentation problems, the bias of the segmentation label is an affine function of the true label.

APPENDIX D

DETERMINATION OF THE OPTIMUM PARAMETERS FOR THE MODIFIED CRAMÉR–RAO BOUND

We first find the optimum values of K_i and u_i for the modified Cramér–Rao bound (27) by setting the derivative of (27) with respect to K_i and u_i to zero, respectively. Then the modified Cramér–Rao bound is obtained through submitting the resulting K_i^* and u_i^* into (27).

$$
\begin{aligned}
&\frac{\partial CRB_{Biased}^{Mod}(\hat{H}_i)}{\partial u_i} = \frac{\partial}{\partial u_i}\left\{\mathrm{Tr}\left((I+K_i)\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}(I+K_i)^T\right) + \int (K_iH_i+u_i)^T(K_iH_i+u_i)P(H)dH\right\} = 0 \\
&\Rightarrow \int\left\{\frac{\partial}{\partial u_i}(K_iH_i+u_i)^T(K_iH_i+u_i)\right\}P(H)dH = 0 \\
&\Rightarrow \int\{2(K_iH_i+u_i)\}P(H)dH = 0 \\
&\Rightarrow u_i = -K_i\int H_i\,p(H)dH = -K_iE_{H_i}(H_i)
\end{aligned}
\tag{D.1}
$$

Using (D.1), we have

$$
\begin{aligned}
&\frac{\partial CRB_{Biased}^{Mod}(\hat{H}_i)}{\partial K_i} = \frac{\partial}{\partial K_i}\left\{\mathrm{Tr}\left((I+K_i)\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}(I+K_i)^T\right) + \int (K_iH_i+u_i)^T(K_iH_i+u_i)P(H)dH\right\} = 0 \\
&\Rightarrow 2(I+K_i)\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + \int 2(K_iH_i+u_i)H_i^{\,T}P(H)dH = 0 \\
&\Rightarrow K_i\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + \int H_iH_i^{\,T}P(H)dH\right\} = -\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + \int u_iH_i^{\,T}P(H)dH\right\} \\
&\Rightarrow K_i\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + \int H_iH_i^{\,T}P(H)dH\right\} = -\left(\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} - \int_H K_iE_{H_i}(H_i)H^T p(H)dH\right) \\
&\Rightarrow K_i\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + E\left(H_iH_i^{\,T}\right)\right\} = -\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} - K_iE_{H_i}(H_i)E_{H_i}(H_i)^T\right\} \\
&\Rightarrow K_i\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + E\left(H_iH_i^{\,T}\right) - E_{H_i}(H_i)E_{H_i}(H_i)^T\right\} = -\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} \\
&\Rightarrow K_i\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + Cov(H_i)\right\} = -\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} \\
&\Rightarrow K_i^* = -\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + Cov(H_i)\right\}^{-1}
\end{aligned}
\tag{D.2}
$$

So

$$
\begin{aligned}
u_i^* &= -K_iE_{H_i}(H_i) \\
&= \left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} + Cov(H_i)\right\}^{-1}E_{H_i}(H_i)
\end{aligned}
\tag{D.3}
$$

Substituting K_i^* and u_i^* into $CRB_{Biased}^{Mod}(\hat{H}_i)$, we obtain the modified bound for the ith region as follows

$$
\begin{aligned}
CRB_{Biased}^{Mod}(\hat{H}_i)^* &= \mathrm{Tr}\left(\left(I+K_i^*\right)\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left(I+K_i^*\right)^T\right)+\int\left(K_i^*H_i+u_i\right)^T\left(K_i^*H_i+u_i^*\right)P(H)dH \\
&= \mathrm{Tr}\left(\left(I+K_i^*\right)\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left(I+K_i^*\right)^T\right)+\int\left(K_i^*H_i-K_iE_{H_i}(H_i)\right)^T\left(K_i^*H_i-K_iE_{H_i}(H_i)\right)P(H)dH \\
&= \mathrm{Tr}\left(\left(I+K_i^*\right)\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left(I+K_i^*\right)^T\right)+\int\left(H_i-E_{H_i}(H_i)\right)^T K_i^{*T}K_i^*\left(H_i-E_{H_i}(H_i)\right)P(H)dH \\
&= \mathrm{Tr}\left(\left(I+K_i^*\right)\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left(I+K_i^*\right)^T\right)+E_{H_i}\left\{\left(H_i-E_{H_i}(H_i)\right)^T K_i^{*T}K_i^*\left(H_i-E_{H_i}(H_i)\right)\right\} \\
&= \mathrm{Tr}\left(\left(I+K_i^*\right)\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left(I+K_i^*\right)^T\right)+E_{H_i}\left\{\mathrm{Tr}\left[K_i^*\left(H_i-E_{H_i}(H_i)\right)\left(H_i-E_{H_i}(H_i)\right)K_i^{*T}\right]\right\} \\
&= \mathrm{Tr}\left(\left(I+K_i^*\right)\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left(I+K_i^*\right)^T\right)+\mathrm{Tr}\left\{K_i^*E_{H_i}\left[\left(H_i-E_{H_i}(H_i)\right)\left(H_i-E_{H_i}(H_i)\right)\right]K_i^{*T}\right\} \\
&= \mathrm{Tr}\left\{\left(I+K_i^*\right)\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left(I+K_i^*\right)^T+K_i^*E_{H_i}\left[\left(H_i-E_{H_i}(H_i)\right)\left(H_i-E_{H_i}(H_i)\right)\right]K_i^{*T}\right\} \\
&= \mathrm{Tr}\left\{\left(I+K_i^*\right)\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left(I+K_i^*\right)^T+K_i^*Cov_{H_i}(H_i)K_i^{*T}\right\} \\
&= \mathrm{Tr}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}+2K_i^*\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}+K_i^*\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}K_i^{*T}+K_i^*Cov_{H_i}(H_i)K_i^{*T}\right\} \\
&= \mathrm{Tr}\left\{\begin{array}{l}\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}-2\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}+Cov(H_i)\right\}^{-1}\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} \\ +\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}+Cov(H_i)\right\}^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}+Cov_{H_i}(H_i)\right\} \\ \left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}+Cov(H_i)\right\}^{-1}\right\}^T\end{array}\right\} \\
&= \mathrm{Tr}\left\{\begin{array}{l}\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}-2\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}+Cov(H_i)\right\}^{-1}\left(E_\beta\left[J_F(H_i)\right]\right)^{-1} \\ +\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}+Cov(H_i)\right\}^{-1}\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\end{array}\right\} \\
&= \mathrm{Tr}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}-\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\left\{\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}+Cov(H_i)\right\}^{-1}\left(E_\beta\left[J_F(H_i)\right]\right)^{-1}\right\}
\end{aligned}
\tag{D.4}
$$

REFERENCES

[1] Y. Zhang (Editor), *Advances in Image And Video Segmentation*, IRM Press, USA May 2, 2006.

[2] P. K. Sahoo, Soltani, and A. K. C.Wong, "A survey of thresholding techniques," *Computer Vision, Graphics, and Image Processing archive*, vol. 41, no. 2, Feb., 1988.

[3] S. Z. Li, *Markov Random Field Modeling in Computer Vision*. New York: Springer-Verlag, 2001.

[4] Gerhard Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction (2nd edition)*, Springer, Feb., 27, 2006.

[5] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 1549–1560, Apr. 1995.

[6] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functionals and associated variational problems," *Commun. Pure Appl. Math.*, vol. 42, pp. 577–685, 1989.

[7] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern recognition*, vol. 26, no. 9, pp. 1277-1294 1993.

[8] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, May 2002.

[9] Yi Ma, Harm Derksen, Wei Hong and John Wright, "*Segmentation of multivariate mixed data via lossy coding and compression*," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 9, pp. 1546- 1562, 2007.

[10] J. Canny, "A computational approach to edge detection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[11] L. Shapiro and G. Stockman, *Computer Vision*, Prentice-Hall, Inc. 2001.

[12] D. Geman, S. Geman, Chr. Graffigne, and P. Dong, "Boundary detection by constrained optimization," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 609-628, 1990.

[13] D. Geman and S. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721-741, 1984.

[14] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.

[15] D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape," *Int. J. Comput. Vis.*, vol. 72, no. 2, pp. 195–215, 2007.

[16] J. Freixenet, X. Munoz, D. Raba, J. Marti and X. Cuffi, "Yet another survey on image segmentation: region and boundary information integration," in *Proc. of the 7th European Conference on Computer Vision-Part III*, pp. 408 – 422, 2002.

[17] J. Fan, D.K.Y. Yau, A.K. Elmagarmid, and W.G. Aref, "Automatic image segmentation by integrating color-edge extraction and seeded region growing," *IEEE Trans. on Image Processing*, vol. 10, no. 10, pp. 1454 – 1466. 2001.

[18] J. Wu, and A. C. S. Chung, "A segmentation model using compound Markov random fields based on a boundary model," *IEEE Trans. on Image Processing*, vol. 16, no. 1, Jan. 2007.

[19] M. S. Alili and D. Ziou, "An approach for dynamic combination of region and boundary information in segmentation," in *Proc. 19th International Conference on Pattern Recognition*, 2008.

[20] H. Zhang, J. E. Fritts, and S. A. Goldman, "Image segmentation evaluation: a survey of unsupervised methods," *Computer Vision and Image Understanding*, vol. 110, p. 260-280, 2008.

[21] Y. Zhang, "A survey on evaluation methods for image segmentation," *Pattern Recognition*, vol. 29, no. 8, pp. 1335-1346, 1996.

[22] T. Lei, and J. K. Udupa, "Performance evaluation of finite normal mixture model-based image segmentation techniques," *IEEE Trans. on Image Processing*, vol. 12, no. 10, pp. 1153-1169, Oct. 2003.

[23] I. B. Kerfoot, and Y. Bresler, "Theoretical analysis of multi-spectral image segmentation criteria," *IEEE Trans. on*

Image Processing, vol. 8, no. 6, pp. 798-820, Jun. 1999.

[24] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE Trans. on Medical Imaging*, vol. 23, no. 7, pp. 903-921, Jul. 2004.

[25] T. Hastie and R. Tibshirani, "Varying-coefficient models," *Journal of the Royal Statistical Society*, series B, vol. 55, no. 4, pp. 757-796, 1993.

[26] F. Salzenstein, and W. Pieczynski, "Parameter estimation in hidden fuzzy Markov random fields and image segmentation," *Graphical Models and Image Processing*, vol. 59, no. 4, pp. 205–220, Jul., 1997.

[27] B. M. Sadler, R. J. Kozick, and T. Moore, "Bounds on bearing and symbol estimation with side information," *IEEE Trans. on Signal Processing*, vol. 49, no. 4, Apr. 2001.

[28] M. Rivera, O. Ocegueda, and J. L. Marroquin, "Entropy-controlled quadratic Markov measure field models for efficient image segmentation," *IEEE Trans. on Image Processing*, vol. 16, no. 12, pp. 3047-3057, Dec. 2007.

[29] J. L. Marroquin, E. A. Santana, and S. Botello, "Hidden Markov measure feld models for Image segmentation," *IEEE Trans. on Patter Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1380-1387, Nov. 2003.

[30] H. L. V. Trees, and K. L. Bell (ed.), *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, Wiley-IEEE Press, 2007

[31] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, "Some classes of global Cramér–Rao bounds," *Ann. Stat.*, vol. 15, pp. 1421-1438, 1987.

[32] P. Stoica and B.C.Ng, "On the Cramèr-Rao bound under parametric constraints," *IEEE Signal Processing Lett.*, vol. 5,pp. 177-179, Jul., 1998.

[33] D. R. Farenick, and F. Zhou, "Jensen's inequality relative to matrix-valued measures," *Journal of Mathematical Analysis and Applications*, vol. 327, no. 2, pp. 919-929, 2007.

[34] S. Kay and Y.C. Eldar, "Rethinking biased estimation," *Signal Processing Magazine*, vol. 25, no. 3, pp. 133-136, May, 2008.

[35] Y. C. Eldar, "Minimum variance in biased estimation: bounds and asymptotically optimal estimators," *IEEE Trans. Signal* Processing, vol. 52, pp. 1915-1930, July 2004.

[36] Y. C. Eldar, "MSE bounds with affine bias dominating the Cramèr-Rao bound," *IEEE Trans. on Signal Processing*, vol. 56, no. 8, Aug., 2008.

[37] P. Chatterjee and P. Milanfar, "Is denoising dead?" *IEEE Transactions on Image Processing*, vol. 19, num. 4, pp 895-911, Apr. 2010.

[38] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transaction on Image Processing*, vol. 13, pp. 600-612, 2004.

[39] A. C. Davison, and D.Hinkley, *Bootstrap Methods and their Application* (8th ed.), Cambridge: Cambridge Series in Statistical and Probabilistic Mathematics, 2006.

[40] H. J. Adèr, G. J., Mellenbergh, and D. J. Hand, *Advising on research methods: A consultant's companion*, Huizen, The Netherlands: Johannes van Kessel Publishing, 2008.

[41] [online]. Available: http://marathon.csee.usf.edu/Mammography/Database.html.

[42] Gerhard Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction (2nd edition)*, Springer, Feb., 27, 2006.

[43] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. on Sys., Man., Cyber, v*ol. 9, pp.62–66, 1979.

[44] P. Liao, T, Chen, and P. Chung, "A fast algorithm for multilevel thresholding," *J. Inf. Sci. Eng.*, vol. no. 5, pp.713–727, 2001.

[45] P. A. Devijver, and J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice Hall, Englewood Cliffs, London, 1982.

[46] C. Li, C. Kao, J. C. Gore, and Z. Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Trans. Image Processing*, vol. 17, no.10, pp.1940-1949, 2008.

[47] T. Cour, F. Benezit, and J. Shi, "Spectral segmentation with multi-scale graph decomposition," *IEEE* International Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[48] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[49] D. Dubois, and H. Prade, *Fundamentals of Fuzzy Sets*, Boston, Kluwer, 2000.

[50] I. Gath and A.B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 7, pp. 773-781, 1989.

[51] Y. Xia, D. Feng, and R. Zhao, "Adaptive segmentation of textured images by using the coupled Markov random field model," *IEEE Trans. on Image Processing*, vol. 15, no. 11, pp. 3559-3566, 2006.

[52] A. Wiesel, Y. C. Eldar and A. Yeredor, "Linear regression with Gaussian model uncertainty: algorithms and bounds," *IEEE Trans. on Signal Processing*., vol. 56, no. 6, pp. 2194-2205, June 2008.

[53] Y. C. Eldar, "Minimax MSE estimation of deterministic parameters with noise covariance uncertainties," *IEEE Trans. Signal Processing*, vol. 54, no. 1, pp. 138-145, Jan. 2006.

[54] Y. C. Eldar, "Minimax estimation of deterministic parameters in linear models with a random model matrix," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 601-612, Feb. 2006.