Anisotropic Scale Selection, Robust Gaussian [1] Fitting, and Pulmonary Nodule Segmentation in Chest CT Scans

Kazunori Okada

Department of Computer Science

San Francisco State University

San Francisco, CA, 94132, USA

`kazokada@sfsu.edu`

Abstract

This chapter presents the theory and design principles used to derive semi-automatic algorithms for pulmonary nodule segmentation toward realizing a reliable and reproducible clinical application for nodule volumetry. The proposed algorithms are designed to be robust against the variabilities due to 1) user-interactions for algorithm initialization, 2) attached or adjacent non-target structures, and 3) non-standard shape and appearance. The proposed theory offers an elegant framework to introduce the robust data analysis techniques into a solution for nodule segmentation in chest X-ray computed tomography (CT) scans. The theory combines two distinct concepts for generic data analysis: automatic scale selection and robust Gaussian model fitting. The unification is achieved by 1) relating Lindeberg's scale selection theory in Gaussian scale-space [1], [2] to Comaniciu's robust feature space analyses with mean shift in Gaussian kernel density estimation [3], [4] and 2) extending both approaches to consider *anisotropic* scale from their original isotropic formulations. This chapter demonstrates how the resulting novel concept of *anisotropic scale selection* gives a useful and robust solution to the Gaussian fitting problem used as a part of our robust nodule segmentation solutions.

Keywords: segmentation, pulmonary nodules, chest CT, automatic scale selection, anisotropic scale-space, Gaussian scale-space, Gaussian fitting, robust estimation, mean shift, scale-space mean shift.

I. INTRODUCTION

Lung caner is the leading cause of cancer death in the U.S. with an estimated 219,440 new cases expected in 2009 [5]. Chest X-ray computed tomography (CT) scan offers one of the most effective diagnostic tools for this cancer in both primary and metastatic cases. Lung cancer in a CT scan commonly exhibits a focal concentration of high-intensity values inside lung parenchyma, known as pulmonary nodules. Pulmonary nodules vary largely in their geometry, topology, and pathology [6]. Nodules may appear solitary or attached to other pulmonary structures, such as blood vessels and pleural surface [7]. The size of visible nodules varies from 1 to 30 mm in diameter [8]. While many small nodules (e.g., $< 10mm$) are benign, some of them can be malignant, whose correct diagnosis plays a key role for early detection of lung cancers [9]. Nodules can also appear solid, as well as part- or non-solid, known as ground-glass opacity (GGO) nodules [10], [11]. The GGO nodules are clinically significant due to its link to aggressive adenocarcinoma [11], [12] and technically challenging to characterize due to its ambiguous appearance [13], [14], [15].

A. The Problem: Pulmonary Nodule Segmentation

Pulmonary nodule segmentation aims to delineate the extent of these nodules in CT, providing a critical foundation of computer-aided diagnosis (CAD) for lung cancers [16], [17], [18]. Due to its increasing clinical significance, the pulmonary nodule segmentation has been actively studied in recent years along with the rapid advances of the high-resolution thin-slice and multi-detector CT technologies [19], [20], [21], [22], [7], [23], [24], [25], [26], [14], [27], [28], [29], [30], [31], [32], [15], [33], [34], [35].

Accurate nodule segmentation is a crucial prerequisite for many diagnostic and treatment procedures for lung cancer, such as studying tumor growth in follow-up [36], [37], monitoring tumor response to therapy [38], [39], screening for early detection [40], [41], and classifying tumor malignancy [42], [43]. Nodule volumetry, the measurement of 3D volume of a nodule, requires accurate segmentation [36], [37], [38], [39], [44], [45], [46], [30], [31], [47], [48], [49], [50], [51]. In a tumor follow-up or therapy monitoring study, tumor growth/response can be characterized by differentiating nodule's volume measured at different time-points, replacing the traditional 2D-based RECIST and WHO protocols [38]. The segmentation also defines a local image area, from which image features can be extracted for further computational analyses. For

example, lung cancer screening by computer-aided-detection (CADe) often enhances the overall detection accuracy by segmenting detected nodules as a post-analysis to remove false-positive cases [52]. Tumor malignancy classification in computer-aided diagnosis (CADx) will also rely on accurate segmentation for extracting image appearance features whose quality dictates the overall classification performance [53]. Thus improving accuracy of nodule segmentation has direct impact to these clinical tasks. While segmentation of large solitary nodule can be straightforward, small or GGO nodules cause difficulty because of the partial volume effect (PVE) [13], [15], [30], [31].

B. The Motivation: Reproducibility in Nodule Volumetry

Beyond the demand for accuracy, the major technical challenge facing the pulmonary nodule segmentation is reproducibility. Recent reports on nodule volumetry have revealed that there are significant inter-scan and inter-observer variability in volume measurement by using the segmentation solutions currently available at the clinical practice [39], [44], [46], [47], [48], [49], [50]. This limits clinical applications of the segmentation-based volumetry to characterize tumor's growth in a short time interval. The causes for the inter-scan variability include variations in image acquisition and reconstruction settings [44], [46], [49], while those for inter-observer variability include uncertainty in the results of segmentation algorithm employed [47], [48], [50]. Pulmonary nodule segmentation is a semi-automatic procedure, involving user-determined seed points to indicate a target nodule to be segmented. Different readers, or a single reader studying the same scan more than once, may produce different seed points, causing different segmentation results of the same nodule. In percentage error of estimated volume, this inter-observer variability can be as high as 20% [39]. Despite the increasing interests in pulmonary nodule segmentation, there is a general lack of studies that focus on designing a reliable and robust segmentation solution against such variabilities. This chapter contains materials previously published by the author in [24], [25], [14]. The rest of chapter is organized in five sections, each of which is dedicated to describe our overall segmentation algorithm design, the theory of anisotropic scale selection, the robust anisotropic Gaussian fitting algorithms, the results of our experimental evaluations, and our conclusive remarks, respectively.

II. OVERALL SEGMENTATION ALGORITHM DESIGN

We design reproducible nodule segmentation solutions by reducing inter-observer and inter-scan variabilities in their results. Our approach consists of the three successive steps as follows:

1) Fit an anisotropic Gaussian model to an input image containing a target nodule, given a single voxel seed located nearby the target,

2) Extract an ellipsoidal boundary approximation from the fitted Gaussian as a parametric characterization of the target, and

3) Segment the target nodule more accurately by refining the parametric characterization using the fitted ellipsoid as an initialization of a further non-parametric segmentation procedure.

1) Gaussian model fitting: plays a key role for designing our robust nodule segmentation solution. Suppose that the intensity distribution $f(\mathbf{x})$ of nodule's CT appearance can be approximated by a product of a Gaussian function (i.e., multivariate normal distribution) $\Phi(\mathbf{x})$ and a positive amplitude parameter α.

$$g(\mathbf{x}) \equiv \alpha \times \left. \Phi(\mathbf{x}) \right|_{\mathbf{x} \in \mathcal{S}} \tag{1}$$

where $\mathcal{S}$ is a set of data points in the neighborhood of the center location $\mathbf{u}$, in which the Gaussian approximation is supposed to be valid such that

$$f(\mathbf{x}) \approx \left. g(x) \right|_{\mathbf{x} \in \mathcal{S}} \tag{2}$$

A d-variate isotropic Gaussian $\Phi_i(\mathbf{x})$ is defined by

$$\Phi_i(\mathbf{x}; \mathbf{u}, \sigma^2) \equiv (2\pi\sigma^2)^{-d/2} \exp\left(-\frac{(\mathbf{x} - \mathbf{u})^t (\mathbf{x} - \mathbf{u})}{2\sigma^2}\right) \tag{3}$$

Such a Gaussian model $g(\mathbf{x}; \mathbf{u}, \sigma, \alpha)$ with $d = 3$ can be fitted to the nodule's CT appearance by estimating the center $\mathbf{u}$, width σ, and height α of the Gaussian that appears most similar to the input.

In order to regularize this estimation process over discrepancy between the nodule's true appearance and the standard Gaussian model, we consider smoothing image data. Gaussian scale-space [54], [55], [56], [1], [57], [58] provides such smoothed data in a series of images blurred by Gaussian filters of increasing widths, Gaussian scale-space $L : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}$ is a one-parameter family of a d-variate continuous signal $f : \mathbb{R}^d \to \mathbb{R}$ provided by a convolution with isotropic Gaussian kernels $\Phi_i(\mathbf{x}; \mathbf{0}, h)$ with increasing (band)widths or scales $h \geq 0$.

$$L(\mathbf{x}; h) \equiv f(\mathbf{x}) * \Phi_i(\mathbf{x}; \mathbf{0}, h^2) \equiv \int_{\mathbb{R}^d} f(\mathbf{y}) \, \Phi_i(\mathbf{x} - \mathbf{y}; \mathbf{0}, h^2) d\mathbf{y} \tag{4}$$

Such a linear scale-space is known to be a solution of the diffusion equation $\partial_h L = 1/2 \nabla^2 L$ [54], [55] initialized by $L(\mathbf{x}; 0) = f(\mathbf{x})$.

Automatic scale selection is an interesting problem defined over this scale-space:

Find the scale from a set of analysis scales (i.e., filter widths) that provides the best estimate of the local image structure's size, known as characteristic scale.

A well-known solution to this problem was first proposed by Lindeberg [1], [2]; The characteristic scale h^* of a local image structure is defined by the local maximum of the normalized scale-space derivatives over scales h and space $\mathbf{x}$. When blob-like nodular structures are considered, the derived characteristic scale h^* and its location x^* can be treated as the width σ and center $\mathbf{u}$ of the Gaussian fitted to the underlying blob. In this sense, Lindeberg's scale selection solves the Gaussian fitting problem. This concept with a Laplacian of Gaussian (LoG) case has recently been exploited for pulmonary nodule segmentation [59], [35]. However, the major shortcoming of this approach is that the isotropic Gaussian approximation is too restrictive to capture the true appearance of nodules accurately. In the 3D domain, this amounts to approximating a nodule by a spherical blob. An obvious extension is to employ an *ellipsoid* to approximate a nodule by fitting an *anisotropic* Gaussian $\Phi_a(\mathbf{x})$,

$$\Phi_a(\mathbf{x}; \mathbf{u}, \boldsymbol{\Sigma}) \equiv |2\pi\boldsymbol{\Sigma}|^{-/2} \exp(-\frac{1}{2}(\mathbf{x} - \mathbf{u})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{u})) \tag{5}$$

where $|\cdot|$ denotes the matrix determinant and $\boldsymbol{\Sigma}$ is a $d \times d$ symmetric positive semi-definite (SPSD) covariance matrix, which determines the shape of the d-variate anisotropic Gaussian. This chapter presents two approaches to address this extension.

The first approach exploits the anisotropic scale selection. The Lindeberg's scale selection principle can be extended to anisotropic scale-space $L(\mathbf{x}; \mathbf{H})$ that consists of a group of images blurred by Gaussian filters with varying scales and orientations [1], [60], [25]. Anisotropic scale selection determines the covariance matrix $\boldsymbol{\Sigma}$ by finding a local maximum of the normalized derivatives of the anisotropic scale-space over $d(d+1)/2$ free parameters forming the anisotropic scale $\mathbf{H} \in \mathcal{SPSD}(d)$, where $\mathcal{SPSD}(d)$ denotes a set of all $d \times d$ symmetric positive semi-definite matrices. Various selection criteria have been proposed by Lindeberg, as well as by the author, using the second moment matrix [1] and the Hessian matrix [60], [25]. More details of this approach will be described in Section III.

The second approach is based on a stability-based scale selection inspired by the work by Comaniciu on data-driven bandwidth selection for kernel density estimation (KDE) [3]. In this work, mean shift [4] was used to define a basin of attraction $\mathcal{S}$ of each mode in a KDE function. An isotropic Gaussian was fitted to each mode by robust least-squares estimation of local mean $\mathbf{u}$ and variance σ^2 using data statistics collected only from the basin $\mathcal{S}$. This was iterated for

successively increasing analysis bandwidths, resulting in a set of fitted Gaussian models. The stability-based selection chose the scale that produced the least change in the fitting results when analysis bandwidth was perturbed. To apply this for anisotropic Gaussian fitting, we have translated this method to 1) Gaussian scale-space by exploiting its similarity to the KDE with Gaussian kernel [24], [25] and to 2) anisotropic Gaussian by using a total least-squares with the SPSD constraint [61], [62], [24]. More details of this approach will be described in Section IV.

2) Ellipsoidal boundary approximation: is given by an equal-elevation contour of the fitted anisotropic Gaussian model. Another definition of this boundary is by a point set with equal Mahalanobis distance from the center $\mathbf{u}$ of the fitted Gaussian,

$$d_M(\mathbf{x}; \mathbf{u}, \mathbf{\Sigma}) \equiv (\mathbf{x} - \mathbf{u})^t \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{u}) \tag{6}$$

In 3D, this will form an ellipsoid whose principal axes corresponds to the eigen vectors of the inverse of the estimated covariance $\mathbf{\Sigma}$ and whose radii to the square root of their eigen values. Deriving an ellipsoid from a Gaussian function requires a threshold parameter of the fixed elevation or Mahalanobis distance. In our previous work [14], we have experimentally chosen this distance value as 1.6416, corresponding to the 35% confidence limit of the normal distribution in 3D.

3) Boundary refinement: takes the derived ellipsoidal boundary $\{\mathbf{x} | d_M(\mathbf{x}; \mathbf{u}, \mathbf{\Sigma}) = 1.6416\}$ as an initialization to further refine the boundary segmentation. Accurate volumetry requires this process as the ellipsoidal approximation may cause a significant bias to its volume estimate. Any deformable surface methods, such as the level set method [63], can be used for this task. Other approaches include spatiotemporal mean shift clustering [26] and likelihood ratio test [27].

This chapter focuses on the first of this three-step nodule segmentation framework. The two proposed Gaussian fitting approaches will be revisited in the next sections with their detailed descriptions, as well as discussion on their robustness against variability typical in our data set.

III. ANISOTROPIC SCALE SELECTION

This section presents the theory for anisotropic scale selection. The automatic scale selection relates the characteristic scale of underlying object to the analysis scale used for constructing the scale-space. Extending this to anisotropic scales yields a paradigm to estimate the Gaussian covariance matrix by evaluating data represented in an anisotropic scale-space. It was Lindeberg and Garding who pioneered this concept, calling it affine Gaussian scale-space and affine shape

adaptation [1], [60]. Later Okada et al. [25] extended this framework as anisotropic scale selection to include more scale selection criteria. This section follows the notations used in Okada et al.

A. Anisotropic Scale-Space and Its Derivatives

Anisotropic scale-space is a generalization of the isotropic scale-space in Eq.(4) by considering an anisotropic Gaussian kernel characterized by a fully parameterized symmetric positive semi-definite *analysis scale matrix* $\mathbf{H} \in \mathcal{SPSD}(d) \in \mathbb{R}^{d \times d}$. Anisotropic scale-space $L : \mathbb{R}^d \times \mathbb{R}^{d \times d} \to \mathbb{R}$ of a d-variate continuous signal $f(\mathbf{x})$ is defined as:

$$L(\mathbf{x}; \mathbf{H}) \equiv f(\mathbf{x}) * \Phi_a(\mathbf{x}; \mathbf{0}, \mathbf{H}) \equiv \int_{\mathbb{R}^d} f(\mathbf{y}) \Phi_a(\mathbf{x} - \mathbf{y}; \mathbf{0}, \mathbf{H}) d\mathbf{y} \tag{7}$$

The analysis matrix $\mathbf{H}$ has $d(d + 1)/2$ free parameters and controls the shape of the Gaussian kernel. The anisotropic scale-space is a solution to the anisotropic homogeneous diffusion:

$$\partial_{\mathbf{H}} L(\mathbf{x}; \mathbf{H}) = 1/2 \nabla \nabla^t L(\mathbf{x}; \mathbf{H}) \tag{8}$$

$$L(\mathbf{x}, \mathbf{0}) = f(\mathbf{x}) \tag{9}$$

The nth-order derivatives of $L(\mathbf{x}; \mathbf{H})$ can be derived by convolving the signal $f(\mathbf{x})$ with the nth-order Gaussian derivative kernels since the differential operators commute across the convolution operations. Thus the first-order scale-space gradient vector $G(\mathbf{x}; \mathbf{H}) \in \mathbb{R}^d$ and the second-order scale-space Hessian matrix $\mathbf{\Psi}(\mathbf{x}; \mathbf{H}) \in \mathbb{R}^{d \times d}$ are defined by,

$$G(\mathbf{x}; \mathbf{H}) \equiv \nabla L(\mathbf{x}; \mathbf{H}) = f(\mathbf{x}) * \Phi(\mathbf{x}; \mathbf{H}) \mathbf{H}^{-1}(-\mathbf{x}) \tag{10}$$

$$\mathbf{\Psi}(\mathbf{x}; \mathbf{H}) \equiv \nabla \nabla^t L(\mathbf{x}; \mathbf{H}) = f(\mathbf{x}) * \Phi(\mathbf{x}; \mathbf{H}) \mathbf{H}^{-1}(\mathbf{x}\mathbf{x}^t - \mathbf{H}) \mathbf{H}^{-1} \tag{11}$$

Next we introduce *L-normalized scale-space derivatives* defined by the point-wise division of the scale-space derivatives by the corresponding scale-space as response-normalized derivatives. *L*-normalized scale-space gradient vector G_l and Hessian matrix $\mathbf{\Psi}_l$ are defined by,

$$G_l(\mathbf{x}; \mathbf{H}) \equiv \frac{G(\mathbf{x}; \mathbf{H})}{L(\mathbf{x}; \mathbf{H})} = \frac{f(\mathbf{x}) * \Phi(\mathbf{x}; \mathbf{H}) \mathbf{H}^{-1}(-\mathbf{x})}{f(\mathbf{x}) * \Phi(\mathbf{x}; \mathbf{H})} \tag{12}$$

$$\mathbf{\Psi}_l(\mathbf{x}; \mathbf{H}) \equiv \frac{\mathbf{\Psi}(\mathbf{x}; \mathbf{H})}{L(\mathbf{x}; \mathbf{H})} = \frac{f(\mathbf{x}) * \Phi(\mathbf{x}; \mathbf{H}) \mathbf{H}^{-1}(\mathbf{x}\mathbf{x}^t - \mathbf{H}) \mathbf{H}^{-1}}{f(\mathbf{x}) * \Phi(\mathbf{x}; \mathbf{H})} \tag{13}$$

B. Scale-Space Derivatives of Gaussian-Like Structures

Now we consider an analysis of local blob-like structures using the above anisotropic scale-space derivatives. Suppose that the signal $f(\mathbf{x})$ represents a volume-of-interest (VOI) in a 3D CT image, containing a blob-like structure (e.g., nodule), and that the blob's CT appearance can locally be approximated by the Gaussian model $g(\mathbf{x}; \mathbf{u}, \boldsymbol{\Sigma}, \alpha)$ in Eq.(1) with the anisotropic Gaussian $\Phi_a(\mathbf{x})$ in Eq.(5),

$$f(\mathbf{x}) \approx g(\mathbf{x}; \mathbf{u}, \boldsymbol{\Sigma}, \alpha) = \alpha \times \Phi_a(\mathbf{x}; \mathbf{u}, \boldsymbol{\Sigma})|_{\mathbf{x} \in \mathcal{S}} \tag{14}$$

Note that the above modeling assumes the signal to be positive-valued. Although a CT scan typically comes in the Hounsfield unit that can take negative values, it is straightforward to transform the intensity range so that this positive-value constraint is readily met. With this assumption, the anisotropic scale-space in Eq.(7) takes a form of another Gaussian with a covariance $\boldsymbol{\Sigma} + \mathbf{H}$ since a convolution of two Gaussians is another Gaussian. Similarly, the anisotropic scale-space derivatives in Eqs.(10,11) become a convolution of a Gaussian with Gaussian derivatives, resulting in,

$$L(\mathbf{x}; \mathbf{H}) \simeq \alpha \Phi_a(\mathbf{x}; \mathbf{u}, \boldsymbol{\Sigma} + \mathbf{H}) \tag{15}$$

$$G(\mathbf{x}; \mathbf{H}) \simeq \alpha \Phi_a(\mathbf{x}; \mathbf{u}, \boldsymbol{\Sigma} + \mathbf{H})(\boldsymbol{\Sigma} + \mathbf{H})^{-1}(\mathbf{u} - \mathbf{x}) \tag{16}$$

$$\boldsymbol{\Psi}(\mathbf{x}; \mathbf{H}) \simeq \alpha \Phi_a(\mathbf{x}; \mathbf{u}, \boldsymbol{\Sigma} + \mathbf{H})(\boldsymbol{\Sigma} + \mathbf{H})^{-1}[(\mathbf{u} - \mathbf{x})(\mathbf{u} - \mathbf{x})^t$$
$$- (\boldsymbol{\Sigma} + \mathbf{H})](\boldsymbol{\Sigma} + \mathbf{H})^{-1} \tag{17}$$

Plugging Eqs.(15-17) to Eqs.(12,13) yields two equations that are satisfied when the L-normalized scale-space derivatives are evaluated for a local blob-like structure,

$$G_l(\mathbf{x}; \mathbf{H}) \simeq (\boldsymbol{\Sigma} + \mathbf{H})^{-1}(\mathbf{u} - \mathbf{x}) \tag{18}$$

$$\boldsymbol{\Psi}_l(\mathbf{x}; \mathbf{H}) \simeq (\boldsymbol{\Sigma} + \mathbf{H})^{-1}(\mathbf{u} - \mathbf{x})(\mathbf{u} - \mathbf{x})^t(\boldsymbol{\Sigma} + \mathbf{H})^{-1} - (\boldsymbol{\Sigma} + \mathbf{H})^{-1} \tag{19}$$

Notice that this response-normalization removes both the multiplicative parameter α and the Gaussian term Φ_a from the derivative formulae. Both L-normalized scale-space gradient G_l and Hessian $\boldsymbol{\Psi}_l$ are computable since $L(\mathbf{x}; \mathbf{H})$ is non-zero within a finite range as long as $f(\mathbf{x})$ is positive-valued.

C. Maximum-Over-Scales Criteria for Scale Selection

The maximum-over-scales criterion was first proposed by Lindeberg [2] as a solution to the isotropic scale selection problem. Scale-space derivatives normalized by the analysis scale h raised to the power of an upper-bounded real value γ are defined as γ-*normalized derivatives*. Lindeberg demonstrated that various functions of the γ-normalized derivatives assume their local maximum at the characteristic scale of the target feature. For the d-variate anisotropic Gaussian structures, the γ-normalized Laplacian with $\gamma = (d+2)/4$ evaluated at a spatial local maximum, $\text{tr}[\mathbf{H}^{(d+2)/4}\mathbf{\Psi}(\mathbf{u};\mathbf{H})]$, is locally maximized over scales when the analysis scale $\mathbf{H}$ is equal to the signal's covariance $\mathbf{\Sigma}$, where "tr" denotes the trace of a $d \times d$ matrix. Lindeberg and Garding [60] have shown that this criterion can be extended to other second-order cases using the trace and determinant of second moment and Hessian matrices.

We developed novel anisotropic maximum-over-scales criteria constructed with the *norm* of the combined γ- and L-normalized scale-space derivatives [25]. The norm is defined both in vector and matrix while the trace and determinant used in the Lindeberg's criteria are specific to matrix. Therefore the choice of the norm makes it possible to consider first-order scale selection criteria and to unify them with the second-order criteria using the same functional form. Note also that values of γ in the Lindeberg's scale selection criteria depend on the dimension of the signal and the order of the differentiation for assuming the maximum-over-scales property. The proposed criteria provide an elegant solution, in which a constant $\gamma = 1/2$ gives rise to the maximum-over-scales property regardless of the signal's dimensions and of the order of differentiation. The spatial local maximum location $\mathbf{u}$ is assumed to be known hereafter. Such maximum can be found by searching locations that satisfies $\nabla L(\mathbf{x};\mathbf{H}) = 0$. For notational simplicity, the function arguments of $G_l(\mathbf{x};\mathbf{H})$ and $\mathbf{\Psi}_l(\mathbf{x};\mathbf{H})$ are omitted unless they are evaluated at a specific location. Fig.1 illustrates the proposed criteria with a synthetic 1D Gaussian signal.

1) First-Order Criterion: Using Eq.(18), a γ-normalization of the L-normalized scale-space gradient vector with $\gamma=1/2$ is expressed by

$$\mathbf{H}^{1/2}G_l = \mathbf{H}^{1/2}(\mathbf{\Sigma} + \mathbf{H})^{-1}(\mathbf{u} - \mathbf{x}) \tag{20}$$

where $\mathbf{H}^{1/2}$ denotes the square root of the matrix $\mathbf{H}$: $\mathbf{H} = \mathbf{H}^{1/2}\mathbf{H}^{1/2}$. This matrix square root can be computed readily since $\mathbf{H} \in \mathcal{SPSD}(d)$. The L_2 norm of the γ- and L-normalized gradient is then given by,

$$\|\mathbf{H}^{1/2}G_l\|_2 = \|\mathbf{H}^{1/2}(\mathbf{\Sigma} + \mathbf{H})^{-1}(\mathbf{u} - \mathbf{x})\|_2 \tag{21}$$
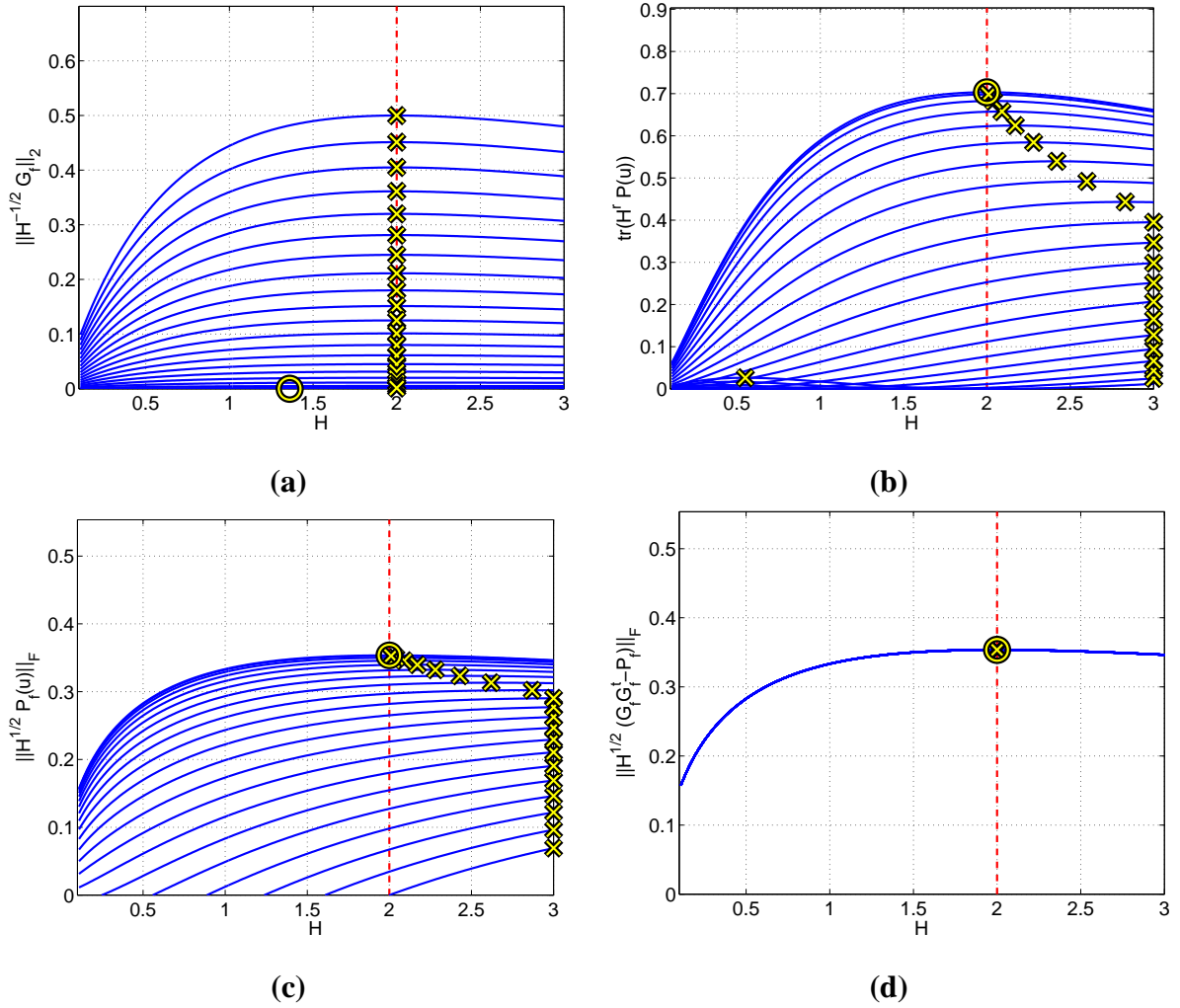
Fig. 1. Examples of the maximum-over-scales criteria. A centered 1D Gaussian signal with $\sigma^2 = 2$ is used as target. (a) the first-order method with Eq.(21), (b) the γ-normalized Laplacian in [2], (c) the second-order method with Eq.(24), (d) the second-order method with Eq.(28). Curved lines in each plot display the norm computed at 21 locations $\mathbf{x}$=0, 0.1, ..., 2 over 291 analysis scales h=0.1, 0.11, .., 3. Dash lines denote the ground-truth scale. "◯" and "×" indicate the maximum-over-scales for the spatial maximum ($\mathbf{x} = 0$) and for the non-maximum ($\mathbf{x} \neq 0$), respectively.

Using this, the following first-order maximum-over-scales criterion is obtained:

Thm1: The L_2 norm of the γ- and L-normalized scale-space gradient vector G_l with $\gamma = 1/2$ in Eq.(21) is maximized when the analysis scale matrix $\mathbf{H} \in \mathcal{SPSD}(d)$ is equal to the characteristic covariance $\boldsymbol{\Sigma}$ of a local Gaussian-like structure and when $\mathbf{x} \neq \mathbf{u}$.

Proof: Comaniciu [3] proved a theorem stating that *the magnitude of the bandwidth-normalized mean shift vector with a Gaussian kernel is maximized when the analysis bandwidth* $\mathbf{H}$ *is equal to* $\boldsymbol{\Sigma}$ [3, p.287]. A proof of this theorem cites the Comaniciu's proof by

showing that the L-normalized scale-space gradient G_l is proportional to the mean shift vector defined over a scale-space: $\mathbf{m}(\mathbf{x};\mathbf{H}) \equiv \mathbf{H}G_l$ [24]. This immediately leads to $\|\mathbf{H}^{1/2}G_l\|_2 = \|\mathbf{H}^{-1/2}\mathbf{m}\|_2$ stating that the L_2 norm of the γ- and L-normalized scale-space gradient vector is equivalent to the magnitude of the bandwidth-normalized mean shift vector. Details of scale-space and mean shift will be revisited in Section IV. When $\mathbf{x} = \mathbf{u}$, $G_l = 0$ for all $\mathbf{H}$, thus the norm does not hold the maximum-over-scales property. $\square$

Note that this criterion holds at arbitrary locations $\mathbf{x} \in \mathcal{S}$ except at the center $\mathbf{u}$ as demonstrated in Fig.1(a).

2) Second-Order Criteria: Two types of second-order scale selection criteria are derived using the Frobenius matrix norm, which is a common extension of the L_2 norm to a matrix. First, a solution only with the second-order Hessian matrix is examined. Using Eq.(19), the γ- and L-normalized scale-space Hessian matrix with $\gamma=1/2$ is expressed by,

$$\mathbf{H}^{1/2}\mathbf{\Psi}_l = \mathbf{H}^{1/2}G_lG_l^t - \mathbf{H}^{1/2}(\mathbf{\Sigma} + \mathbf{H})^{-1} \tag{22}$$

When evaluated at the spatial maximum $\mathbf{u}$, the normalized Hessian is reduced to the following simple form because G_l becomes zero,

$$\mathbf{H}^{1/2}\mathbf{\Psi}_l(\mathbf{u};\mathbf{H}) = -\mathbf{H}^{1/2}(\mathbf{\Sigma} + \mathbf{H})^{-1} \tag{23}$$

The Frobenius matrix norm of this derivative matrix function is then given by,

$$\|\mathbf{H}^{1/2}\mathbf{\Psi}_l(\mathbf{u};\mathbf{H})\|_F = \|\mathbf{H}^{1/2}(\mathbf{\Sigma} + \mathbf{H})^{-1}\|_F \tag{24}$$

The following second-order maximum-over-scales criterion is obtained using Eq. (24),

Thm2: The Frobenius norm of the γ- and L-normalized scale-space Hessian matrix with $\gamma = 1/2$ in Eq.(24) is maximized when the analysis scale matrix $\mathbf{H} \in \mathcal{SPSD}(d)$ is equal to $\mathbf{\Sigma}$ and when $\mathbf{x} = \mathbf{u}$.

Proof: We define $\eta(\mathbf{H}) \equiv \|\mathbf{H}^{1/2}\mathbf{\Psi}_l(\mathbf{u};\mathbf{H})\|_F$. The theorem must be true if $\eta(\mathbf{\Sigma})^2 - \eta(\mathbf{H})^2$ is greater or equal to zero with equality *iff* $\mathbf{H} = \mathbf{\Sigma}$. Recall that $\mathbf{\Sigma}$ and $\mathbf{H}$ are symmetric

positive definite matrices. Thus we have,

$$\eta(\mathbf{\Sigma})^2 - \eta(\mathbf{H})^2$$
$$= \|\mathbf{\Sigma}^{1/2}(\mathbf{\Sigma} + \mathbf{\Sigma})^{-1}\|_F^2 - \|\mathbf{H}^{1/2}(\mathbf{\Sigma} + \mathbf{H})^{-1}\|_F^2$$
$$= \tfrac{1}{4}\text{tr}[\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2}] - \text{tr}[(\mathbf{\Sigma} + \mathbf{H})^{-1}\mathbf{H}(\mathbf{\Sigma} + \mathbf{H})^{-1}]$$
$$= \tfrac{1}{4}\text{tr}[\mathbf{\Sigma}^{-1} - 4(\mathbf{\Sigma} + \mathbf{H})^{-1}\mathbf{H}(\mathbf{\Sigma} + \mathbf{H})^{-1}]$$
$$= \tfrac{1}{4}\text{tr}[(\mathbf{\Sigma} + \mathbf{H})^{-1}(\mathbf{H}\mathbf{\Sigma}^{-1} - \mathbf{I})^2\mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{H})^{-1}]$$

Since $\mathbf{\Sigma}$ and $\mathbf{H}$ are positive definite, all the matrices inside the trace are also positive definite. Since the trace of a positive definite matrix is positive-valued, we have $\eta(\mathbf{\Sigma})^2 - \eta(\mathbf{H})^2 \geq 0$. Trivially, the equality holds iff $\mathbf{H} = \mathbf{\Sigma}$. □

The classic d-variate γ-normalized Laplacian by Lindeberg can be expressed as a matrix trace,

$$\text{tr}[\mathbf{H}^{(d+2)/4}\mathbf{\Psi}(\mathbf{u}; \mathbf{H})] = -L(\mathbf{u}; \mathbf{H})\text{tr}[\mathbf{H}^{(d+2)/4}(\mathbf{\Sigma} + \mathbf{H})^{-1}] \tag{25}$$

Also the Frobenius norm in Eq.(24) can be expressed by,

$$\|\mathbf{H}^{1/2}\mathbf{\Psi}_l(\mathbf{u}; \mathbf{H})\|_F^2 = \text{tr}[(\mathbf{\Sigma} + \mathbf{H})^{-1}\mathbf{H}(\mathbf{\Sigma} + \mathbf{H})^{-1}] \tag{26}$$

As compared in Fig.1(b) and Fig.1(c), both methods behave similarly despite the difference in their functional forms. Theorem 2 is true only at the spatial maximum $\mathbf{u}$ as shown in Fig.1(c) as was also the case for the Lindeberg's second-order criteria.

Second, both first-order gradient and second-order Hessian is examined together by considering a matrix function of the difference between the L-normalized second-moment matrix $G_l G_l^t$ and Hessian matrix $\mathbf{\Psi}_l$. From Eqs.(18,19), the γ-normalization of this matrix function with $\gamma{=}1/2$ is given by,

$$\mathbf{H}^{1/2}(G_l G_l^t - \mathbf{\Psi}_l) = \mathbf{H}^{1/2}(\mathbf{\Sigma} + \mathbf{H})^{-1} \tag{27}$$

The Frobenius norm of this normalized derivative function is,

$$\|\mathbf{H}^{1/2}(G_l G_l^t - \mathbf{\Psi}_l)\|_F = \|\mathbf{H}^{1/2}(\mathbf{\Sigma} + \mathbf{H})^{-1}\|_F \tag{28}$$

Consequently, we obtain the following second-order maximum-over-scales criterion using Eq.(28),

Thm3: Consider a matrix function subtracting the L-normalized scale-space Hessian matrix from the L-normalized second moment matrix: the outer-product of the L-normalized scale-space gradient vector. The Frobenius norm of the γ-normalization of this matrix function with $\gamma = 1/2$ in Eq.(28) is maximized for all $\mathbf{x} \in \mathcal{S}$ when the analysis scale matrix $\mathbf{H} \in \mathcal{SPSD}(d)$ is equal to $\mathbf{\Sigma}$.

Proof: For all $\mathbf{x} \in \mathcal{S}$, $\|\mathbf{H}^{1/2}(G_l G_l^t - \boldsymbol{\Psi}_l)\|_F = \|\mathbf{H}^{1/2}(\boldsymbol{\Sigma} + \mathbf{H})^{-1}\|_F = \eta(\mathbf{H})$. From the proof[13] of theorem 2, we have $\eta(\boldsymbol{\Sigma})^2 - \eta(\mathbf{H})^2 \geq 0$ with equality at $\mathbf{H} = \boldsymbol{\Sigma}$. $\square$

As shown in Fig.1(d), the function value is *invariant* against the location $\mathbf{x}$; the value does not depend on the location, at which it is evaluated. Note a peculiar equality of two scale space derivative functions: the L-normalized scale-space Hessian evaluated at the spatial maximum ($\boldsymbol{\Psi}_l(\mathbf{u}; \mathbf{H})$) and the difference of the L-normalized second moment and Hessian matrices ($G_l G_l^t - \boldsymbol{\Psi}_l$). Furthermore, unlike the other first- and second-order criteria, the maximum-over-scales property of this criterion holds at any locations $\mathbf{x} \in \mathcal{S}$. This is a key property of this criterion which allows us to collect the scale estimates from a number of points rather than to have a single point-estimate only at the spatial maximum $\mathbf{x} = \mathbf{u}$. This property will be exploited in another scale selection strategy which will be described in the next section.

D. Related Work

The extension of the Gaussian scale-space to anisotropic scales has been explored in past. They have been called in various terms, including anisotropic scale-space [25], affine scale-space [1], [60], and directional scale-space [64]. The anisotropic scale-space discussed in this chapter models the anisotropic homogeneous diffusion process while the traditional linear Gaussian scale-space [54], [55] models the isotropic homogeneous diffusion and the anisotropic diffusion proposed by Perona and Malik [65] models the isotropic inhomogeneous diffusion.

Lindeberg and Garding [1], [60] is the first to discuss the type of anisotropic extension of the Gaussian scale-space discussed in this chapter and applied it to fingerprint analysis [66]. The work presented in this chapter complements Lindeberg's work with unifying scale selection criteria and robust estimation techniques introduced to apply them to practical problems. Introduction of robust estimation techniques to the scale-space has been scarcely studied in past. Robust anisotropic diffusion proposed by Black et al. [67] is the only other related work addressing this issue to the best of our knowledge.

Automatic scale selection was pioneered by Lindeberg for isotropic scale [2] and later for anisotropic scale [1], [60], [25]. This scale selection concept has played a major role in developing the popular scale- and affine-invariant interest-point detector such as SIFT [68], Harris-Hessian Affine feature detectors [69], and GRIF [64]. The proposed second-order criterion in Eq.(28) has an interesting parallel to the affine-invariant feature detector proposed by Mikolajczyk and

Schmid [69]. The Harris-affine and Hessian-affine detectors by Mikolajczyk and Schmid are<superscript>14</superscript> based respectively on the second moment matrix and Hessian matrix, while the proposed second-order criterion combines both second moment and Hessian matrices in a single scale selection method.

The anisotropic extension for the Gaussian kernel has also been explored in the context of convolution filter design in the orientation space. Freeman and Adelson proposed the oriented steerable filter by a linear combination of basis filters [70], while Bigun et al. [71] presented multi-dimensional directional estimation using the matrix eigenvalue problem. These classic oriented filters have also been extended to 3D domain. Faas and van Vliet [72] discussed an extension of the steerable filter to the 3D orientation space with pseudo-regular 3D orientation sampling using the icosahedron grid. On the other hands, Lampert and Wirjadi [73] proposed a method for separating an anisotropic Gaussian filter along arbitrary oriented axes.

A number of previous works have also addressed applications of the Gaussian intensity model fitting using the anisotropic scale-space analysis. Bahalerao and Wilson [74] proposed an intensity model fitting method for visualizing 3D vascular structures in MR images. They utilized anisotropic model similar to ours, however their method does not exploit the robust statistics and involves an expensive EM algorithm-based iterative solution for the model fitting. Our proposed method exploits a more efficient closed-form robust least-squares method. Manmatha and Srimal [75] developed a hand writing segmentation system utilizing an anisotropic scale-space-based blob detection technique. However, they considered anisotropy only up to a diagonal covariance matrix. The extension to the full anisotropy along arbitrary axis is provided by our solutions.

IV.. ROBUST ANISOTROPIC GAUSSIAN FITTING

Let us now revisit the problem of how to robustly fit an anisotropic Gaussian model $g(\mathbf{x})$ in Eq.(1) to the appearance of pulmonary nodules shown in a 3D CT image. One simple and common approach is to treat the image $f(\mathbf{x})$ as a 3D histogram then estimate the Gaussian parameters as the expected mean and covariance matrix with the histogram. The underlying probability distribution required to compute the expectations is calculated by normalizing the histogram. This is possible since $f(\mathbf{x})$ is assumed to be non-negative-valued. This naive method fails when:

1) the intensity distribution in the image $f(\mathbf{x})$ does not follow a Gaussian distribution closely,

2) the target nodule is attached or adjacent to other non-target structures, and

3) the user-specified voxel seed given to indicate a target nodule is placed too far from the target.

The first cause of failure, *non-Gaussianity*, is relevant to some extent for all nodule cases because there is no justification for nodules to have a Gaussian appearance. Moreover, the GGO and necrosis (i.e., cell deaths causing a hollow area in the middle of a nodule) cases will cause larger deviation between the nodule appearance and Gaussian distribution. The second cause of failure, *juxtaposed non-target structure*, makes the image data fit to a Gaussian distribution only within an unknown and irregular range of the image domain. Pulmonary nodules are commonly attached to blood vessels and pleural surface (i.e., lung wall). Because the degree of attachment varies largely from peripheral to submersion, the domain range of which the data follows the Gaussian appearance varies widely as a consequence. Technically speaking, this introduces an arbitrary truncation to a data model, making the model fitting difficult. Even when the target nodule is not attached to other structures, there can be a number of other structures located nearby. A natural extension of the Gaussian fitting approach would be to employ a Gaussian mixture model which can be fit to the multi-structured data by using the Expectation-Maximization (EM) algorithm [76]. However, this approach is not suitable for our problem because many structures do not follow the Gaussian appearance assumption (e.g., vessel) and the number of nearby structures is arbitrary and unknown. The third cause of failure, *seed variability*, is a fundamental limitation of a semi-automatic segmentation and the main cause of the inter-observer variability in volumetry. Together with the non-Gaussianity of data, the Gaussian centroid estimate can yield biased results depending on where the seed is placed. Due to the non-target structures located nearby, the model can be fit to a wrong structure when the seed is located too close to the non-target structure.

Addressing these issues, we propose a semi-automatic approach for fitting a single anisotropic Gaussian model. We designed our model fitting algorithm to be robust against the three afore-mentioned factors common in nodule segmentation: non-Gaussianity, juxtaposed non-target structures, and seed variability. We first derived a convergent scale-space mode seeking algorithms, *scale-space mean shift* [77], [24], by exploiting the similarity of Gaussian scale-space and kernel density estimation. The results of the anisotropic scale selection described in the previous section will be reused with this scale-space mean shift to design a Gaussian fitting algorithm that is

robust against the juxtaposed non-target structures and the seed variability. To address the non-Gaussianity issue, we propose a novel scale selection method, *most-stable-over-scales*. This method carries out the model fitting across the isotropic scale-space then chooses the best fitted model using a stability-based criterion.

The robustness plays a key role toward the goal of realizing a reproducible volumetry. The Gaussian approximation of the nodule appearance poses a trade-off between accuracy and robustness. The approximation causes unavoidable errors in volume measurement of nodules. On the other hand, the mathematical simplicity of the model allows us to robustify the volume estimation process so that the tumor growth rate estimated by such robust methods can be more accurate than the rate estimated by more accurate but less robust segmentation methods, as discussed in [23]. In the following, the proposed robust Gaussian model fitting algorithms will be described.

A. Mean Shift and Anisotropic Scale-Space

Mean shift is a popular technique for analyzing the mode structure of kernel density estimation (KDE) from discrete data samples. It is an adaptive gradient-ascent algorithm with automatic step-size selection and is provably convergent to a mode of the density function. This data-driven framework provides an efficient solution to the general non-parametric data clustering problem in the Parzen windows setting [78], [79], [4]. The mode-seeking property of the mean shift algorithm has also been successfully applied to a wide range of vision problems such as tracking [80], [77] and segmentation [81], [4], [14].

Given a d-variate sample set $\mathbb{S} = \{\mathbf{x}_i \in \mathbb{R}^d | i = 1, .., n\}$, the KDE $\hat{f}(\mathbf{x})$ with a Gaussian kernel $\Phi_a(\mathbf{x})$ in Eq.(5) is defined by,

$$\hat{f}(\mathbf{x}) \equiv \frac{1}{A} \sum_{i=1}^{n} \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H}) \tag{29}$$

where A is a normalizing factor that is a function of n and $\mathbf{H}$. Mean shift vector using the Gaussian kernel is defined as the difference between the kernel-weighted sample mean and the current location,

$$\mathbf{m}(\mathbf{x}) \equiv \frac{\sum_{i=1}^{n} \mathbf{x}_i \, \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H})}{\sum_{i=1}^{n} \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H})} - \mathbf{x} \tag{30}$$

The gradient-ascent like mean shift procedure is then defined by a successive iteration $\mathbf{y}_k$ of this mean shift vector $\mathbf{m}(\mathbf{y}_k)$ from an initialization $\mathbf{y}_0$ until convergence at $\mathbf{y}^*$,

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \mathbf{m}(\mathbf{y}_k) \tag{31}$$

The mean shift formula $\mathbf{m}(\mathbf{x})$ can be derived by evaluating an estimator of the KDE's[17] spatial gradient,

$$
\begin{aligned}
\nabla_x \hat{f}(\mathbf{x}) &= \frac{1}{A} \sum_{i=1}^{n} \nabla_x \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H}) \\
&= \frac{\mathbf{H}^{-1}}{A} \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{x}) \, \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H}) \\
&= \frac{\mathbf{H}^{-1}}{A} \left[\sum_{i=1}^{n} \mathbf{x}_i \, \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H}) - \mathbf{x} \sum_{i=1}^{n} \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H}) \right] \\
&= \frac{\mathbf{H}^{-1}}{A} \sum_{i=1}^{n} \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H}) \times \left[\frac{\sum_{i=1}^{n} \mathbf{x}_i \, \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H})}{\sum_{i=1}^{n} \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H})} - \mathbf{x} \right] \\
&= \mathbf{H}^{-1} \hat{f}(\mathbf{x}) \mathbf{m}(\mathbf{x})
\end{aligned}
\tag{32}
$$

This formulation also clarifies the relationship of the mean shift vector to the KDE's gradient as

$$
\mathbf{m}(\mathbf{x}) = \frac{\mathbf{H} \nabla_x \hat{f}(\mathbf{x})}{\hat{f}(\mathbf{x})}
\tag{33}
$$

Beyond the KDE, the mean shift vector can also be derived from a broad class of kernel-smoothed functions [82], including the scale-space [77], [24]. A naive discretization of the Gaussian scale-space in Eq.(7) yields,

$$
\hat{L}(\mathbf{x}; \mathbf{H}) \equiv \sum_{i=1}^{n} f(\mathbf{x}_i) \, \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H})
\tag{34}
$$

where n is the number of voxels in the image $f(\mathbf{x})$. Given that the image $f(\mathbf{x})$ is positive-valued, this discretized Gaussian scale-space can be considered as a kernel-smoothed function with positive weights after appropriate normalization. Cheng [79] has shown that mean shift derived from such a weighted KDE holds the desirable properties of the original mean shift. Evaluating the gradient of the discretized Gaussian scale-space in Eq.(34) in the same way demonstrated in Eq.(32) yields the scale-space mean shift,

$$
\mathbf{m}(\mathbf{x}) \equiv \frac{\sum_{i=1}^{n} \mathbf{x}_i f(\mathbf{x}_i) \, \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H})}{\sum_{i=1}^{n} f(\mathbf{x}_i) \, \Phi_a(\mathbf{x}; \mathbf{x}_i, \mathbf{H})} - \mathbf{x}
\tag{35}
$$

Applying this to the same means shift procedure in Eq.(31) provides a gradient-ascent like iterative algorithm that is convergent to an intensity peak in $\hat{L}(\mathbf{x}; \mathbf{H})$ in the neighborhood of an arbitrary initialization.

B. Most-Stable-Over-Scales Methods for Gaussian Fitting

As mentioned in Sec II, we use the Gaussian scale-space to regularize an image to handle data irregularity. In constructing a scale-space, the image $f(\mathbf{x})$ is smoothed by Gaussian kernels, modeling the continuous diffusion process. By increasing the kernel bandwidth, an initially noisy image would increasingly become smoother. An important question in this set-up is when to stop this diffusion. With too little smoothing, we have high variance in estimation due to the remaining noise. With too much smoothing, we have a biased estimate causing inaccuracy but with a low variance. This trade-off is known as *bias-variance dilemma* and motivates the problem of *automatic bandwidth selection* in finding most accurate KDE of data samples.

The proposed anisotropic scale selection criteria discussed in Sec III offers solutions to this bandwidth selection problem because of the mathematical duality between the KDE and the scale-space we have established in Sec IV-A. However, this approach poses two serious shortcomings. First, seeking the scale-space maxima becomes impractical when high-dimensional anisotropic structures are considered. Such cases require a dense sampling of a multivariate product space, resulting in prohibitively large search space, known as the *course of dimensionality*. Second, what the maximum-over-scales criteria select as the best scale is not the scale that best regularizes the signal but rather the scale that best matches the target's characteristics scale. When these two scales, the characteristic and best regularizer scales are different, the anisotropic scale selection in Sec III would stop the diffusion at the wrong place.

Addressing these issues, we developed the *most-stable-over-scales* method: a stability-based scale selection that *decouples* the above two types of scales.In order to address the curse of dimensionality in searching the anisotropic scale-space, we employ the isotropic scale-space in Eq.(4) for regularizing the estimation process with linear scales $\mathbf{H} = h\mathbf{I}$ ($h \in \mathbb{R} > 0$). At each analysis scale h, the anisotropic characteristic scale $\boldsymbol{\Sigma}(h)$ is estimated by robustly fitting the Gaussian model $g(\mathbf{x})$ to $f(\mathbf{x})$ by using the scale-space mean shift. The usage of the isotropic scale-space in solving the anisotropic scale selection problem is only possible because of this decoupling. Given a set of S analysis scales $\{h_s | s = 1, .., S, h_s < h_{s+1}\}$, this process yields a set of Gaussian estimates $\{\varPhi_s(\mathbf{x}; \mathbf{u}(h_s), \boldsymbol{\Sigma}(h_s)) | s = 1, .., S\}$.

The most-stable-over-scales scale selection method chooses the best Gaussian $\varPhi^*(\mathbf{x}; \mathbf{u}^*, \boldsymbol{\Sigma}^*)$ among the set $\{\varPhi_s\}$ by selecting the best scale h^* with a divergence-based stability test:

$$(\mathbf{u}^*, \boldsymbol{\Sigma}^*) = (\mathbf{u}(h^*), \boldsymbol{\Sigma}(h^*) \mid h^* = \text{argmin}_s \text{div}\{\varPhi(h_i(s))\}) \tag{36}$$

At each analysis scale h_s, the stability test perturbs the scale within a fixed interval $h_s \pm a$ and fits a Gaussian at every perturbed scale $\{h_i(s)|i = 1, .., L\}$ where $h_1(s) = h_s - a$ and $h_L(s) = h_s + a$. A divergence measure of these Gaussians represents how stable the model fitting process is at the analysis scale h_s. The minimum of this divergence measure over all evaluated analysis scales indicates the most stable Gaussian estimate. A form of the Jensen-Shannon divergence for normal distributions proposed in [3] is employed as our measure, given a set of ordered analysis scales forming a geometric sequence $\{h_s|s = 1, .., S; h_{s+1}/h_s = \text{constant}\}$,

$$
\begin{aligned}
JS(h_s) &= \frac{1}{2} \log \frac{|\frac{1}{2a+1}\sum_i \boldsymbol{\Sigma}(h_i(s))|}{\sqrt[2a+1]{\prod_i |\boldsymbol{\Sigma}(h_i(s))|}} \\
&+ \frac{1}{2}\sum_i (\mathbf{u}(h_i(s)) - \mathbf{u})^t (\sum_i \boldsymbol{\Sigma}(h_i(s)))^{-1}(\mathbf{u}(h_i(s)) - \mathbf{u})
\end{aligned}
\tag{37}
$$

where $\mathbf{u} = \frac{1}{2a+1}\sum_i \mathbf{u}(h_i(s))$. Jensen-Shannon divergence is a natural extension of the pair-wise Kullback-Leibler divergence to describe similarity among a set of distributions [83]. In practice, the index i and the interval a can be set so that the stability test at each scale h_s evaluates with the three immediately neighboring scales (h_{s-1}, h_s, h_{s+1}).

The proposed automatic scale/bandwidth selection method is robust against the non-Gaussianity of the nodule data because it does not analyze the residual error of the model fitting, which is a common bandwidth selection approach. Such residual errors fail to indicate the goodness of fit when the data do not follow the Gaussian appearance assumption.

Robust Gaussian fitting is performed as two successive steps of estimating the Gaussian mean and covariance at each analysis scale h_s. Addressing the seed variability issue, we use majority voting of the convergence points of the scale-space mean shift initialized by the seed. Addressing the juxtaposed non-target structures, we propose a robust least-squares estimation of the covariance matrix by a data-driven estimate of the domain range $\mathcal{S}$ in our model equation Eq.(14). We define *basin of attraction* as a set of all data points in the image domain $\mathbf{x}$, from which the scale-space mean shift in Eq.(35) converges to the same point,

$$
\mathcal{S} \equiv \{\mathbf{x}_j | \forall_j \mathbf{y}_c \approx \mathbf{y}^*(\mathbf{x}_j)\}
\tag{38}
$$

where $\mathbf{y}^*(\mathbf{x}_j)$ denotes a convergence of a scale-space means shift procedure initialized by $\mathbf{x}_j$. This least-squares estimation is robust because the parameters are estimated only from statistics sampled within this basin. The following two sections describe these two estimation processes in details.

C. Mean Estimation

Suppose that the user-specified seed marker $\mathbf{x}_p$ is given inside the basin $\mathcal{S}$. We perform a uniform sampling in the neighborhood of $\mathbf{x}_p$, resulting in N_1 sample points. Using each sample as an initialization, the scale-space means shift is performed on $L(\mathbf{x}; h_s)$, yielding N_1 convergence points $\{\mathbf{y}_j^*\}$. These convergence points are subjected to a clustering with a fixed Mahalanobis distance threshold then the location with the majority vote will provide the final estimate of the Gaussian centroid $\mathbf{u}(h_s)$.

Robustness against the seed variability is important not only for reducing the inter-observer variability but also for enhancing the usability of the overall semi-automatic segmentation by relieving the user from the task of locating the seeds precisely at the nodule's center. The voting scheme will increase the robustness especially when the seed is placed at a peripheral area of the basin. The bias-variance dilemma suffers this estimation task. With too little smoothing, a nodule can exhibit multiple peaks with zero gradient, resulting in high variance in the convergence point set $\{y_j^*\}$. On the other hand, too much smoothing can assure a better-defined basin of attraction with a single peak but with less accurate localization. The task of finding the best estimate is delegated to the most-stable-over-scales method in Eqs.(36,37).

D. Covariance Estimation

Recall the anisotropic scale selection criteria discussed in Sec III. While the Lindeberg's original Laplacian-based criterion is only valid at the spatial maximum location $\mathbf{u}$, the proposed first-order and second-order criteria in Eqs.(21,28) are valid at any other locations in the neighborhood of $\mathbf{u}(h_s)$. Exploiting this property, we estimate the Gaussian covariance $\mathbf{\Sigma}(h_s)$ from $L(\mathbf{x}; h_s)$ by first collecting a set of local linear constraints with the unknown covariance matrix and next solving the set of linear equations to derive the final estimate that minimizes the error in the least-squares sense. Robustness against the juxtaposed non-target structures is achieved by collecting local estimates only from the points within the basin of attraction $\mathcal{S}$ of $\mathbf{u}(h_s)$ so the information that belongs to other non-target structures will be effectively ignored.

To this end, we perform another uniform sampling of data points around the estimated centroid $\mathbf{u}(h_s)$, resulting in N_2 seeds, each of which is used to initialize a scale-space mean shift procedure. At each iteration of a mean shift procedure, a linear constraint equation is computed and stored. The constraint equations collected along the path of the mean shift that converged to $\mathbf{u}(h_s)$ are retained while ignoring those by mean shift converged to elsewhere.

This framework provides a *robust least-squares* method, where the suppression of outliers is given through the basin whose shape is determined from data and adapts to each case unlike the common robust estimators, such as the m-estimator [84]. Note also that our method does not fit the Gaussian model directly to the scale-space $L(\mathbf{x}; h_s)$ but estimates the true covariance $\mathbf{\Sigma}(h_s)$ with $L(\mathbf{x}; h_s)$. This is a key property since it allows a direct comparison of the covariance estimates from different bandwidth h_s in the proposed most-stable-over-scales. In the following, we describe the first- and second-order methods with different linear constraint equations derived from the L-normalized scale-space derivatives in Eqs.(18,19).

1) First-Order Method: A local covariance estimator with the normalized gradient G_l is derived by manipulating Eq.(18) while maintaining its equality,

$$\mathbf{\Sigma} G_l = \mathbf{u} - \mathbf{x} - \mathbf{H} G_l \tag{39}$$

where G_l can be computed numerically by using Eqs.(7,10). This equation can also be expressed as a function of the fixed-bandwidth mean shift vector $\mathbf{m}(\mathbf{x}; \mathbf{H}) \equiv \mathbf{H} G_l(\mathbf{x}; \mathbf{H})$ proposed in [3], [24], i.e., $\mathbf{\Sigma} \mathbf{H}^{-1} \mathbf{m} = \mathbf{u} - \mathbf{x} - \mathbf{m}$. Both equations become singular when G_l goes to zero at $\mathbf{x} = \mathbf{u}$.

The resulting equation of an unknown $\mathbf{\Sigma}$ is under-complete, requiring at least two independent samples for the unique solution. Note that Eq.(39) satisfies with arbitrary $\mathbf{H}$, allowing to use the isotropic scale where $\mathbf{H} = h\mathbf{I}$. Given a sufficient number of independent samples, an over-complete set of linear equations can be formed and solved by a least-squares method with a constraint that the solution must be a symmetric positive semi-definite matrix $\mathbf{\Sigma} \in \mathcal{SPSD}(d)$. We adapt a closed-form constraint least-squares solution proposed in [24] to solve this problem.

Given a set of K measurements $\{(\mathbf{x}_k, G_l(\mathbf{x}_k; h_s)) | k = 1, .., K\}$ within $\mathcal{S}$ along the paths of mean shift procedures convergent to $\mathbf{u}(h_s)$, we construct an over-complete normal equation,

$$\mathbf{A}\mathbf{\Sigma} = \mathbf{B} \tag{40}$$

$$\mathbf{A} \equiv (G_{l1}, .., G_{lK})^t \tag{41}$$

$$\mathbf{B} \equiv (\mathbf{u} - \mathbf{x}_1 - h G_{l1}, .., \mathbf{u} - \mathbf{x}_K - h G_{lK})^t \tag{42}$$

The constrained least-squares solution of this normal equation for the unknown $\mathbf{\Sigma} \in \mathcal{SPSD}(d)$ is given by finding the minimizer $\mathbf{Y}^*$ of an area criterion $\|\mathbf{A}\mathbf{Y} - \mathbf{B}\mathbf{Y}^{-t}\|_F^2$ where $\mathbf{Y}$ is Cholesky factorization of $\mathbf{\Sigma} = \mathbf{Y}\mathbf{Y}^t$ [62]. The closed-form of this solution is expressed by a function of

symmetric Schur decompositions of $\mathbf{\Psi} \equiv \mathbf{A}^t\mathbf{A}$ and $\tilde{\mathbf{Q}} \equiv \mathbf{\Sigma}_P\mathbf{u}_P^t\mathbf{Q}\mathbf{u}_P\mathbf{\Sigma}_P$ given $\mathbf{Q} \equiv \mathbf{B}^t\mathbf{B}$,

$$\mathbf{\Sigma}(h_s) = \mathbf{u}_P\mathbf{\Sigma}_P^{-1}\mathbf{u}_{\tilde{Q}}\mathbf{\Sigma}_{\tilde{Q}}\mathbf{u}_{\tilde{Q}}^t\mathbf{\Sigma}_P^{-1}\mathbf{u}_P^t$$
$$\mathbf{\Psi} = \mathbf{u}_P\mathbf{\Sigma}_P^2\mathbf{u}_P^t \qquad (43)$$
$$\tilde{\mathbf{Q}} = \mathbf{u}_{\tilde{Q}}\mathbf{\Sigma}_{\tilde{Q}}^2\mathbf{u}_{\tilde{Q}}^t$$

2) Second-Order Method: Another covariance estimator with the normalized Hessian $\mathbf{\Psi}_l$ is derived by manipulating Eq.(19) while maintaining its equality,

$$\mathbf{\Sigma} = (G_lG_l^t - \mathbf{\Psi}_l)^{-1} - \mathbf{H} \qquad (44)$$

where G_l can be computed numerically by using Eqs.(7,10) and $\mathbf{\Psi}_l$ by using Eqs.(7,11). This equation exploits both first- and second-order derivatives. Unlike the first-order equation, the equality holds at arbitrary locations $\mathbf{x} \in \mathcal{S}$.

At the spatial maximum $\mathbf{u}$, Eq.(44) collapses into the form only with the Hessian matrix,

$$\begin{aligned} \mathbf{\Sigma} &= (-\mathbf{\Psi}_l(\mathbf{u};\mathbf{H}))^{-1} - \mathbf{H} \\ &= L(\mathbf{u};\mathbf{H})(-\mathbf{\Psi}(\mathbf{u};\mathbf{H}))^{-1} - \mathbf{H} \end{aligned} \qquad (45)$$

The resulting form is similar to the well-known Hessian-based covariance estimator [85], except the second negative term included due to its scale-space nature. Note that, for the second-order case, the magnitude parameter α can be expressed analytically. The analytical form of the scale-space Hessian matrix evaluated at the spatial maximum $\mathbf{u}$ is given by: $\mathbf{\Psi}(\mathbf{u};\mathbf{H}) = -\alpha(2\pi)^{-d/2}|\mathbf{\Sigma}+\mathbf{H}|^{-1/2}(\mathbf{\Sigma}+\mathbf{H})^{-1}$. This equation can be solved for $\mathbf{\Sigma}$ since $\mathbf{H}+\mathbf{\Sigma} \in \mathcal{SPSD}$, i.e., $\mathbf{\Sigma} = \alpha^{\frac{2}{d+2}}|2\pi(-\mathbf{\Psi}(\mathbf{u};\mathbf{H}))^{-1}|^{-\frac{1}{d+2}}(-\mathbf{\Psi}(\mathbf{u};\mathbf{H}))^{-1} - \mathbf{H}$. Since this and Eq.(45) must be equivalent, the following formula is obtained after some algebra,

$$\alpha = \sqrt{|2\pi(-\mathbf{\Psi}(\mathbf{u};\mathbf{H}))^{-1}|L(\mathbf{u};\mathbf{H})^{d+2}} \qquad (46)$$

The scale-space Hessian $\mathbf{\Psi}(\mathbf{x};\mathbf{H})$ is symmetric *negative* semi-definite if $\mathbf{x}$ is at a stable critical point of $-L(\mathbf{x};\mathbf{H})$. When $\mathbf{\Psi}(\mathbf{u};\mathbf{H})$ is numerically computed by using Eq.(11), it must be assured that $\mathbf{u}$ satisfies this condition so that the estimated $\mathbf{\Sigma}$ by Eq.(45) satisfies the positive definite constraint and Eq.(46) remains as real-valued.

Similar to the first-order method, a set of K measurements $\{(G_l(\mathbf{x}_k;h_s), \mathbf{\Psi}_l(\mathbf{x}_k;h_s))|k = 1,..,K\}$ are sampled within $\mathcal{S}$ along the paths of convergent mean shift procedures. A robust least-squares covariance estimator is given by averaging these local estimates,

$$\mathbf{\Sigma}(h_s) = \frac{1}{K}\sum_{k=1}^{K}\{(G_l(\mathbf{x}_k;h_s)G_l(\mathbf{x}_k;h_s)^t - \mathbf{\Psi}_l(\mathbf{x}_k;h_s))^{-1} - h_s\mathbf{I}\} \qquad (47)$$
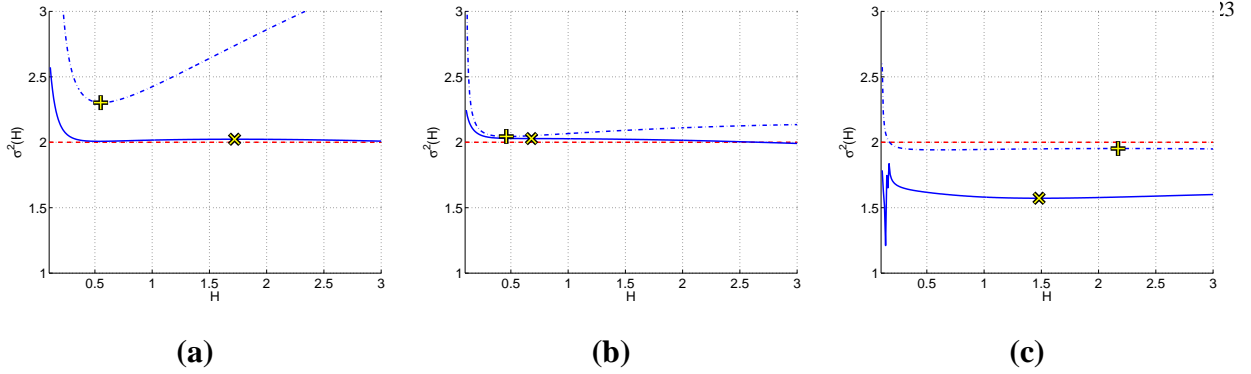
Fig. 2. Examples of the most-stable-over-scales methods. A centered 1D Gaussian signal of $\sigma^2 = 2$ with additive random noise ($randn * 0.01$) is used as the target. At each analysis scale, the variance of the target is estimated from a set of samples within: (a) $\pm0.1\sigma$, (b) $\pm1.0\sigma$, and (c) $\pm3.0\sigma$. Dash lines: the ground-truth scale. Dot lines: the first-order estimates by Eq.(39). Solid lines: the second-order estimates by Eq.(44). "+" and "×" denote the most stable estimates by the first- and second-order methods, respectively.

The second-order equation provides a full covariance estimate for each sample location. Thus a valid estimator with a single sample at the spatial local maximum location $\mathbf{u}$ can be obtained by using Eq.(45).

Fig.2 compares the first- and second-order most-stable-over-scales methods with the 1D synthetic Gaussian data with additive random noise. Three different sampling ranges ($\pm0.1\sigma$, $\pm1.0\sigma$, $\pm3.0\sigma$) were evaluated. Both methods achieve accurate scale estimation given an appropriate choice of the sampling range. The results also suggest that the first-order method favors a larger sampling range while the second-order method prefers a smaller one. When using the data without the noise, both methods resulted in estimates with no errors. The additive noise causes estimation errors that depend on sample locations. For the second-order case, the estimates tend to be more accurate when evaluated at a location closer to the spatial maximum.

V. Experimental Evaluations

A. Synthetic Data with Noise

The proposed scale selection methods are studied with 1D synthetic data with the presence of noises. The target feature is the centered 1D Gaussian with $\sigma^2 = 2$. As shown in Fig.3, three types of additive noise are used: (a) neighboring structure, (b) strong random noise, (c) the combination of (a) and (b).

Fig.4 illustrates the results by the proposed maximum-over-scales criterion. In general, we find that (i) the maximum-over-scales criterion is susceptible to the noises, (ii) the first-order
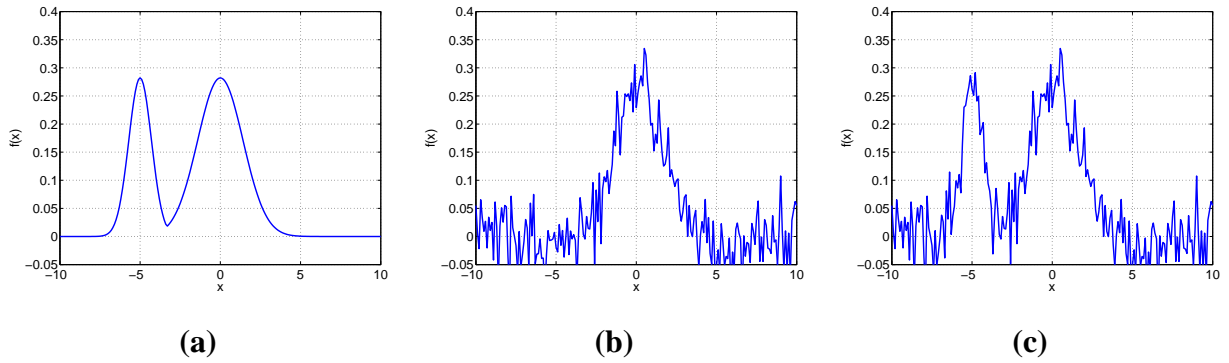
Fig. 3. 1D synthetic data with noise. The target is the centered 1D Gaussian with $\sigma^2 = 2$. (a) a Gaussian centered at $\mathbf{u} = -5$ with $\sigma^2 = 0.5$ superimposed to the target. (b) the target with additive random noise ($randn * 0.04$). (c) the data (a) with the same additive noise in (b).
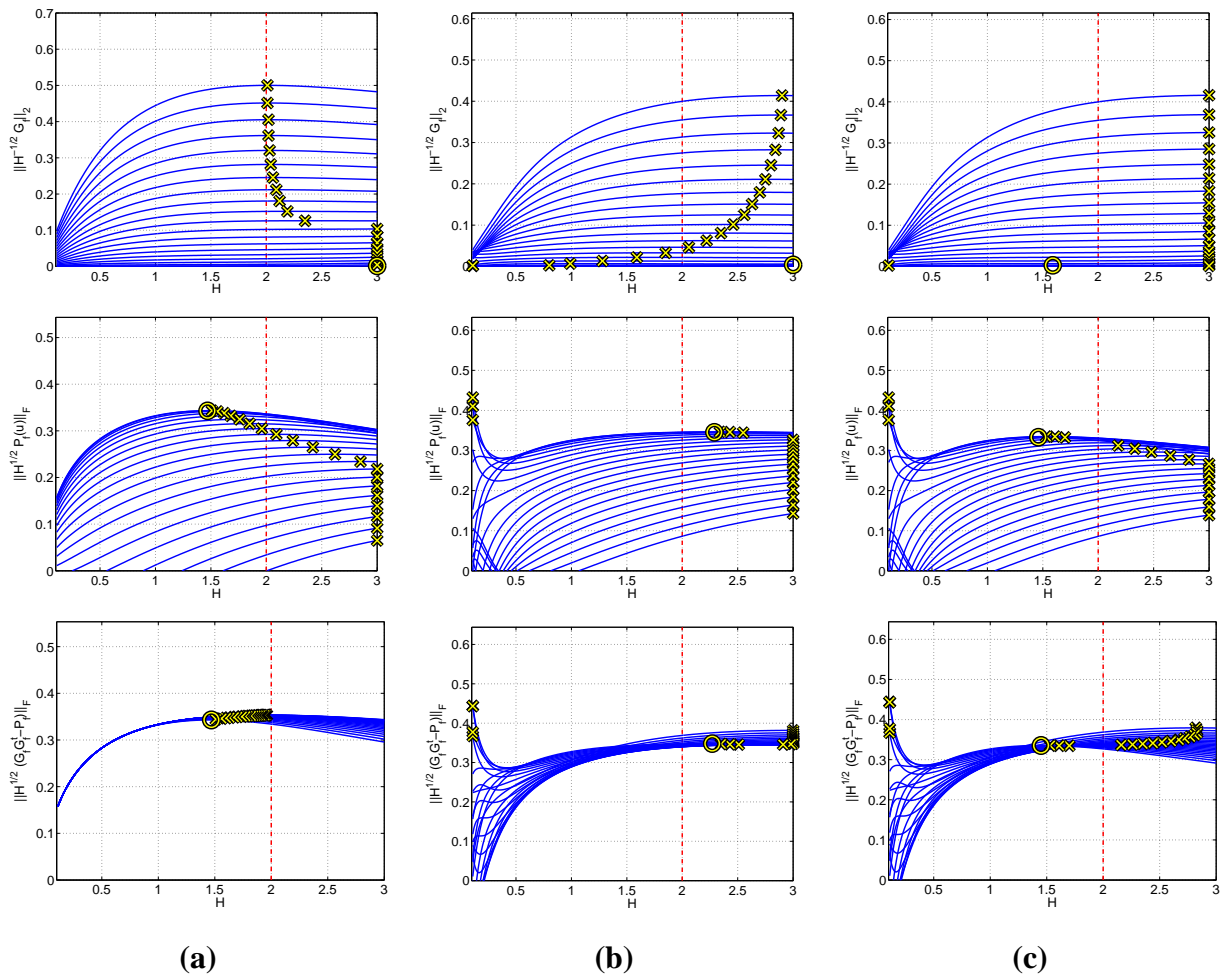

Fig. 4. Variance estimation by the maximum-over-scales criterion for the signals shown in Fig.3. 1st row: the first-order method with Eq.(21). 2nd row: the second-order method with Eq.(24). 3rd row: the second-order method with Eq.(28). The legend is the same as Fig.1.
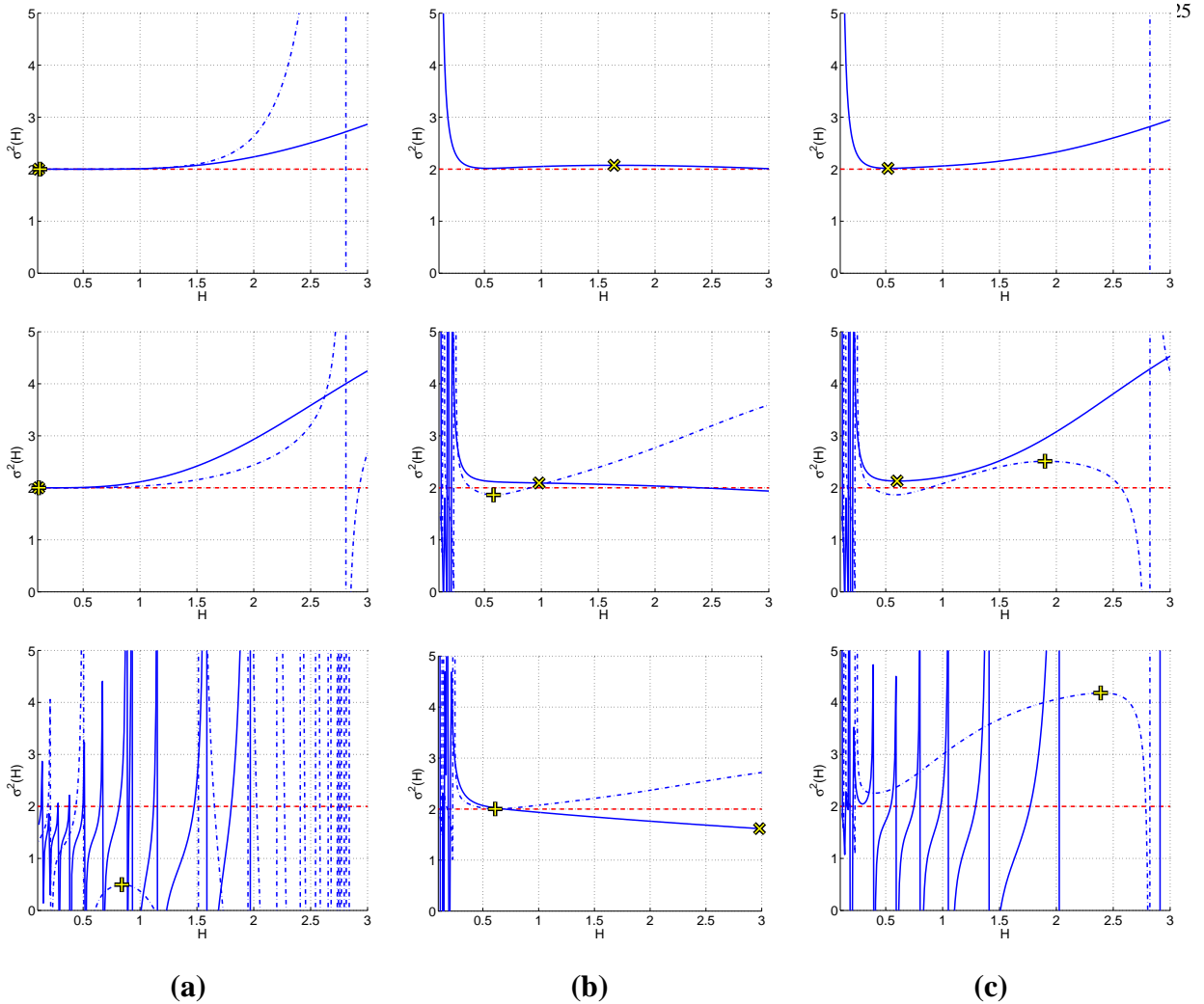
Fig. 5. Variance estimation by the most-stable-over-scales methods for the signals in Fig.3. At each analysis scale, the variance of the target is estimated from samples within: 1st row: $\pm 0.1\sigma$, 2nd row: $\pm 1.0\sigma$, 3rd row: $\pm 2.0\sigma$. The legend is the same as Fig.2.

method is more sensitive to the random noise, (iii) the second-order methods are more sensitive to the neighboring structure. These observations can be explained by the fact that the support of the Gaussian derivative kernels is larger for the higher order derivatives. Thus the second-order methods are naturally more sensitive to the neighboring structure or the signal truncation than the first-order method. The most accurate estimate was obtained by the first-order method when the data without the random noise were evaluated at points far from the non-target structure, as shown in the top-left of Fig.4.

Fig.5 illustrates the results by the most-stable-over-scales criterion. The first-order (dotted lines) and second-order (solid lines) methods are compared by using the same data as Fig.4.
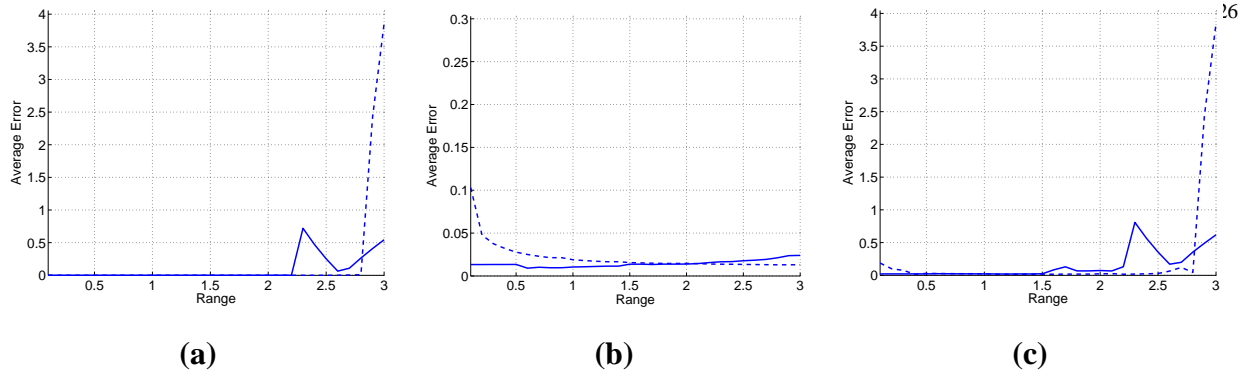
Fig. 6. Average variance estimation errors of the most-stable-over-scales methods over 100 independent tests. The same types of data in Fig.3 with different random noise are used for each test. The errors are plotted against varying sampling ranges. Dot and solid lines denote errors by the first- and the second-order methods, respectively.

At each analysis scale, the target's variance is estimated from samples within three different sampling ranges: $\pm 0.1\sigma$, $\pm 1.0\sigma$, and $\pm 2.0\sigma$. The crosses "+" and "×" denote the estimates by the first- and second-order methods, respectively. The results demonstrate that the most-stable-over-scale criterion are more accurate than the maximum-over-scales criterion if the sampling range is chosen correctly. For the data (a), both methods were accurate using only samples within the basin of attraction. For the data (b), the first-order (second-order) method gave better results with a larger (smaller) range. For the data (c), the second-order method with a very small sampling range was most accurate. The first-order estimate in b(1) and the second-order estimates in a(3) and c(3) were out of range. With the large sampling range, the scale estimates for data (a) and (c) were corrupted because of the samples located near the edge of or out of the target's basin of attraction. The second-order method with the very small sampling range resulted in the overall best accuracy across the different types of noise. Fig.6 shows average estimation errors of the most-stable-over-scales methods over 100 tests. The errors are plotted against continuously varying sampling ranges and compared with the aforementioned three data types. It demonstrates that both the first- and second-order methods achieve much higher accuracy than the maximum-over-scales criterion within the $\pm 2.2\sigma$ sampling range that roughly corresponds to the target's basin of attraction. Also observed was a tendency that the first-order (second-order) method is more accurate with a larger (smaller) sampling range.

Fig.7 illustrates 2D examples comparing the proposed scale selection methods. The test data consists of a centered target Gaussian with additive random noise and a neighboring structure as shown in Fig.7(a). Fig.7(b-e) show results with the maximum-over-scales methods. We use a
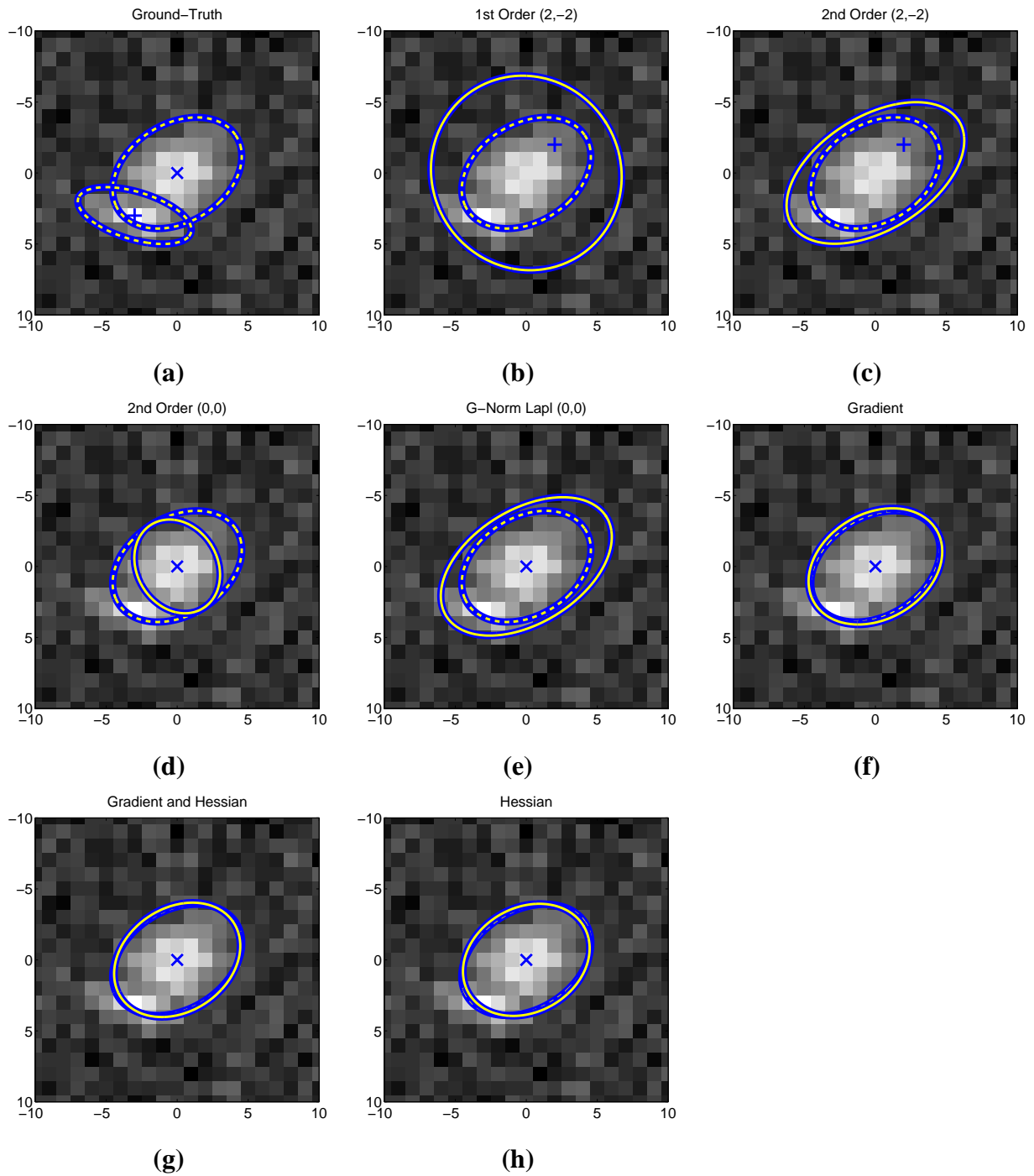
Fig. 7. Examples with 2D synthetic data consisting of a target centered Gaussian with a neighboring structure centered at (-3,3) and additive random noise as shown in (a). From (b) to (e): the maximum-over-scales methods. (b) first-order method evaluated at (2,-2), (c) second-order method at the non-maximum location, (d) second-order method evaluated at (0,0), (e) γ-normalized Laplacian at the maximum location. From (f) to (h): the most-stable-over-scales methods. (f) first-order method, (g) second-order gradient and Hessian method, (h) second-order Hessian only method. The ground-truth and scale estimates are denoted by 90% confidence ellipses with dash and solid lines, respectively.

set of 144 analysis scale matrices sampled along the two eigenvectors of the ground-truth matrix by setting the corresponding eigenvalues to $(1, 2, .., 12)$. Fig.7(f-h) show results with the most-stable-over-scales methods. We used a set of 26 isotropic analysis scales from 0.1 to 7.6 with a constant geometric ratio $2^{1/4}$. The sampling range is set to one Mahalanobis distance. The results suggest that the most-stable-over-scales methods outperform the maximum-over-scales methods, confirming the finding from the 1D case. The three most-stable-over-scales methods resulted in similar accuracy. The second-order case (g) with both gradient and Hessian, however, gave the best accuracy in terms of the Frobenius norm of the error (0.69).

B. Lung CT Data

3D implementations of the most-stable-over-scales methods are applied to the problem of estimating anisotropic spreads of pulmonary nodules shown in high-resolution computed-tomography (HRCT) images. HRCT images of 14 patients displaying the total of 77 pulmonary tumors were used for this evaluation. Each volumetric image consists of 12-bit positive values over an array of 512×512 lattices. A marker indicating the rough tumor location is given by a radiographic reader for each nodule. We compare the first-order and the second-order (Hessian only) methods. For both methods, a set of 14 isotropic analysis scales $h = (0.50^2, .., 4.75^2)$ with a geometric ratio $2^{1/4}$ are used. The system is implemented in C language and process a 32x32x32-voxel volume-of-interest by an average of two seconds with a 2.4 GHz Intel CPU.

The second-order method resulted in less failures (10 cases) than the first-order method (14 cases). All the solitary tumors were correctly estimated by both methods. Most of the failures were due to small nodules that are attached to the lung wall (i.e., on-the-wall cases).

Fig.8 shows examples of the estimation results. The left columns illustrate part- or non-solid nodule cases which are more likely to become malignant than solid ones [10]. The right columns show the cases attached to the pleural surface. Both methods resulted in similar estimates for many cases (e.g., (a)-(e) and (c)-(g)). However, the second-order method often provided more accurate spread estimates (e.g., (b)-(f)). Furthermore, some cases failed by the first-order method were correctly estimated by the second-order method (e.g., (d)-(h)).

VI. Conclusions

This chapter presents a framework for robust pulmonary nodule segmentation, focusing on approximating the nodule's appearance by robustly fitting a Gaussian model to CT data.
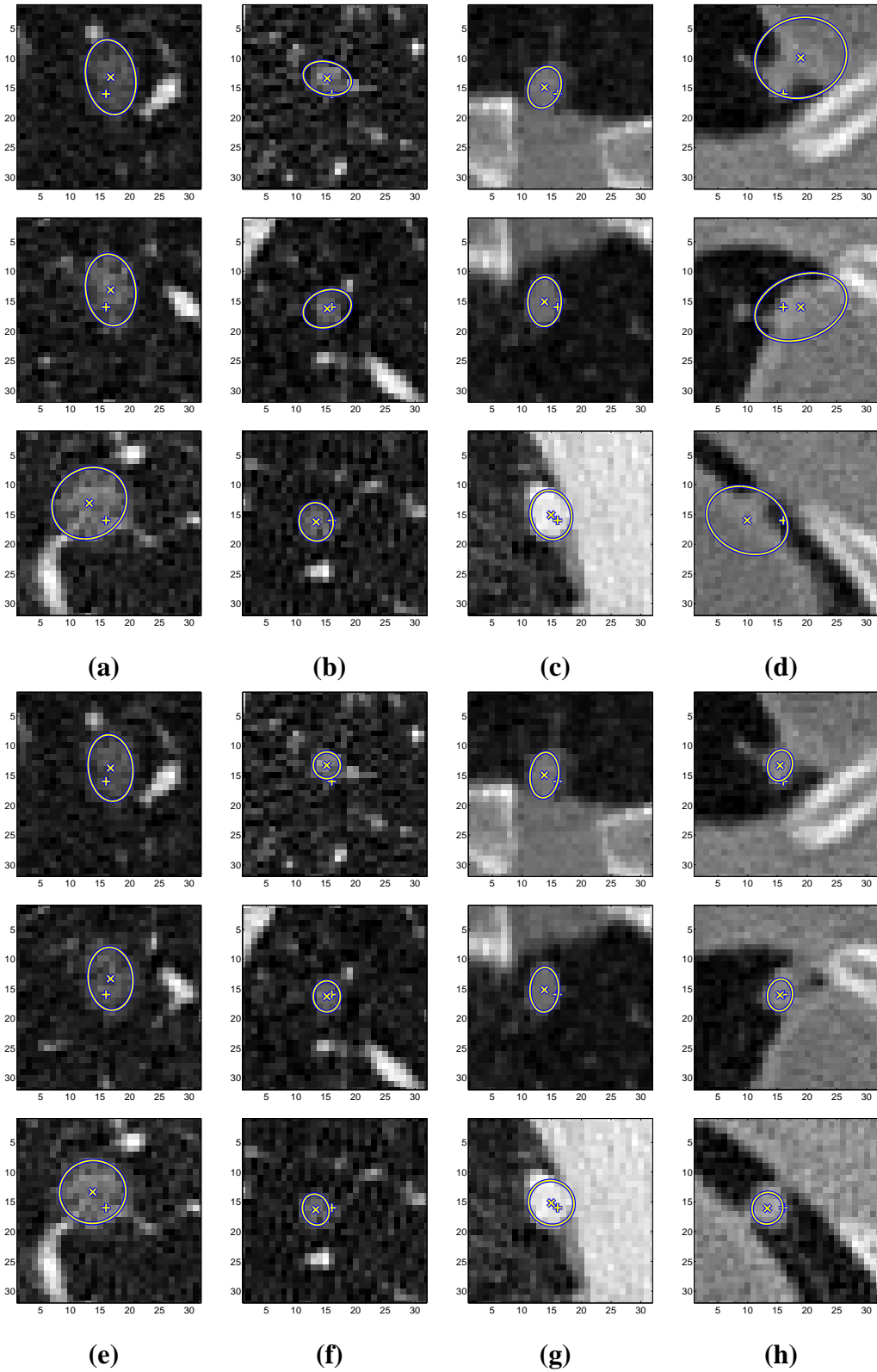
Fig. 8. Examples of the 3D spread estimation of lung tumors in 3D HRCT data by using the most-stable-over-scales scale selection methods. "+" denotes the marker locations. The local spatial maxima and 3D spread estimates are denoted by "×" and 2D intersections of 50% confidence ellipsoids, respectively. Cases (a)-(d) show the results by the first-order method. Cases (e)-(h) show those by the second-order (Hessian only) method.

The framework unifies the concepts of automatic scale selection and robust Gaussian fitting. In particular, we investigated the theory of anisotropic scale selection: scale selection in an anisotropic scale-space. Furthermore, the popular non-parametric mean shift data analysis is extended to the scale-space and we demonstrated the relationship between the automatic scale and bandwidth selection problems. As results, we offer two solutions for anisotropic Gaussian model fitting: 1) anisotropic scale selection with various maximum-over-scales criteria and 2) robust anisotropic Gaussian fitting by various most-stable-over-scales methods.

Technically, the main conclusions of this chapter are (i) the norm of the γ- and L-normalized anisotropic scale-space derivatives offers elegant scale selection solutions with a constant $\gamma=1/2$ regardless of the signal's dimension and of the order of differentiation and (ii) the most-stable-over-scales criterion with scale-space mean shift outperforms the maximum-over-scales criterion in the presence of noise. Experiments with 1D and 2D synthetic data were conducted to validate these findings. Finally, we applied 3D implementations of the proposed methods to the problem of estimating anisotropic spreads of pulmonary nodules shown in HRCT images. Comparison of the first- and second-order methods indicates the advantage of exploiting the second-order information for realistic application scenarios with the presence of noise.

REFERENCES

[1] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.

[2] ——, "Feature detection with automatic scale selection," *Int. J. Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998.

[3] D. Comaniciu, "An algorithm for data-driven bandwidth selection," *IEEE Trans. Patt Anal Mach Intelli*, vol. 25, no. 2, pp. 281–288, 2003.

[4] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Patt Anal Mach Intelli*, vol. 24, no. 5, pp. 603–619, 2002.

[5] A. C. Society, "Cancer facts & figures 2009," Atlanta, 2009.

[6] J. Ko, "Lung nodule detection and characterization with multi-slice CT," *J Thoracic Imag*, vol. 20, pp. 196–209, 2005.

[7] W. Kostis, A. Reeves, D. Yankelevitz, and C. Henschke, "Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images," *IEEE Trans. Medical Imaging*, vol. 22, pp. 1259–1274, 2003.

[8] L. Cardinale, F. Ardissone, S. Novello, M. Busso, F. Solitro, M. Longo, D. Sardo, M. Giors, and C. Fava, "The pulmonary nodule: clinical and radiological characteristics affecting a diagnosis of malignancy," *Radiol med*, vol. 114, pp. 871–889, 2009.

[9] T. Ohtsuka, H. Nomori, H. Horio, T. Naruke, and K. Suemasu, "Radiological examination for peripheral lung cancers and benign nodules less than 10 mm," *Lung Cancer*, vol. 42, pp. 291–296, 2003.

[10] C. Henschke, D. Yankelevitz, R. Mirtcheva, G. McGuinness, and O. McCauley, D. Miettinen, "CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules," *AJR Am. J. Roentgenol.*, vol. 178, no. 5, pp. 1053–1057, 2002.

[11] M. Godoy and D. Naidich, "Subsolid pulmonary nodules and the spectrum of peripheral adenocarcinomas of the lung: recommended interim guidelines for assessment and management," *Radiology*, vol. 253, pp. 606–622, 2009.

[12] J. Min, H. Lee, K. Lee, J. Han, K. Park, M. Ahn, and S. Lee, "Stepwise evolution from a focal pure pulmonary ground-glass opacity nodule into an invasive lung adenocarcinoma: An observation for more than 10 years," *Lung Cancer*, 2010, in press.

[13] I. Sluimer, A. Schilham, P. M., and B. van Ginneken, "Computer analysis of computed tomography scans of the lung: A survey," *IEEE Trans. Medical Imaging*, vol. 25, pp. 385–405, 2006.

[14] K. Okada, D. Comaniciu, and A. Krishnan, "Robust anisotropic Gaussian fitting for volumetric characterization of pulmonary nodules in multislice CT," *IEEE Trans. Medical Imaging*, vol. 24, no. 3, pp. 409–423, 2005.

[15] T. Kubota, A. Jerebko, M. Salganicoff, M. Dewan, and A. Krishnan, "Robust segmentation of pulmonary nodules of various densities: from ground-glass opacities to solid nodules," in *Int Workshop Pulmonary Image Processing*, 2008.

[16] Q. Li, F. Li, K. Suzuki, J. Shiraishi, H. Abe, R. Engelmann, Y. Nie, H. MacMahon, and K. Doi, "Computer-aided diagnosis in thoracic CT," *Seminars in ultrasound, CT, and MRI*, vol. 26, pp. 357–363, 2005.

[17] Q. Li, "Recent progress in computer-aided diagnosis of lung nodules on thin-section CT," *Comput Med Imaging Graph*, vol. 31, pp. 248–257, 2007.

[18] J. Goldin, M. Brown, and I. Petkovska, "Computer-aided diagnosis in lung nodule assessment," *J Thoracic Imag*, vol. 23, pp. 97–104, 2008.

[19] Y. Kawata, N. Niki, H. Ohmatsu, R. Kakinuma, K. Eguchi, M. Kaneko, and N. Moriyama, "Quantitative surface characterization of pulmonary nodules based on thin-section CT images," *IEEE Trans. Nucl Sci*, vol. 45, pp. 2132–2138, 1998.

[20] B. Zhao, A. Reeves, D. Yankelevitz, and C. Henschke, "Three-dimensional multicriterion automatic segmentation of pulmonary nodules of helical computed tomography images," *Optical Engineering*, vol. 38, pp. 1340–1347, 1999.

[21] Y. Lee, T. Hara, H. Fujita, S. Itoh, and T. Ishigaki, "Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique," *IEEE Trans. Medical Imaging*, vol. 20, pp. 595–604, 2001.

[22] M. Brown, M. McNitt-Gray, J. Goldin, R. Suh, and D. Sayre, J.W. Aberle, "Patient-specific models for lung nodule detection and surveillance in CT images," *IEEE Trans. Medical Imaging*, vol. 20, pp. 1242–1250, 2001.

[23] W. Mullally, M. Betke, J. Wang, and J. Ko, "Segmentation of nodules on chest computed tomography for growth assessment," *Med Phys*, vol. 31, pp. 839–848, 2004.

[24] K. Okada, D. Comaniciu, N. Dalal, and A. Krishnan, "A robust algorithm for characterizing anisotropic local structures," in *Proc. Euro. Conf. Computer Vision*, 2004, pp. I:549–561.

[25] K. Okada, D. Comaniciu, and A. Krishnan, "Scale selection for anisotropic scale-space: Application to volumetric tumor characterization," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2004, pp. I:594–601.

[26] ——, "Robust 3D segmentation of pulmonary nodules in multislice CT images," in *Proc. Int. Conf. Medical Imaging Computing and Computer-Assisted Intervention*, 2004, pp. II:881–889.

[27] K. Okada, U. Akdemir, and A. Krishnan, "Blob segmentation using joint space-intensity likelihood ratio test: Application to volumetric tumor characterization," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005, pp. II:437–444.

[28] A. El-Baz, A. Farag, G. Gimel'farb, R. Falk, M. El-Ghar, and T. Eldiasty, "A framework for automatic segmentation of lung nodules from low dose chest CT scans," in *Proc. IARP Int. Conf. Pattern Recognition*, 2006.

[29] A. Farag, A. El-Baz, G. Gimel'farb, R. Falk, M. El-Ghar, and T. Eldiasty, "Appearance models for robust segmentation of pulmonary nodules in 3D LDCT chest images," in *Proc. Int. Conf. Medical Imaging Computing and Computer-Assisted Intervention*, 2006.

[30] A. Reeves, A. Chan, D. Yankelevitz, C. Henschke, B. Kressler, and W. Kostis, "On measuring the change in size of pulmonary nodules," *IEEE Trans. Medical Imaging*, vol. 25, pp. 435–450, 2006.

[31] J. Kuhnigk, V. Dicken, L. Bornemann, A. Bakai, D. Wormanns, S. Krass, and H. Peitgen, "Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans," *IEEE Trans. Medical Imaging*, vol. 25, pp. 417–434, 2006.

[32] J. Wang, R. Engelmann, and Q. Li, "Segmentation of pulmonary nodules in three-dimensional CT images by use of a spiral-scanning technique," *Med Phys*, vol. 34, pp. 4678–4689, 2007.

[33] J. Dehmeshki, H. Amin, M. Valdivieso, and X. Ye, "Segmentation of pulmonary nodules in thoracic CT scans: A region growing approach," *IEEE Trans. Medical Imaging*, vol. 27, pp. 467–480, 2008.

[34] S. Diciotti, G. Picozzi, M. Falchini, M. Mascalchi, N. Villari, and G. Valli, "3-D segmentation algorithm of small lung nodules in spiral CT images," *IEEE Trans. Info Tech Biomed*, vol. 12, pp. 7–19, 2008.

[35] S. Diciotti, S. Lombardo, G. Coppini, L. Grassi, M. Falchini, and M. Mascalchi, "The LoG characteristic scale: A consistent measurement of lung nodule size in CT imaging," *IEEE Trans. Medical Imaging*, vol. 29, pp. 397–409, 2010.

[36] D. Yankelevitz, A. Reeves, W. Kostis, B. Zhao, and C. Henschke, "Small pulmonary nodules: Volumetrically determined growth rates based on CT evaluation," *Radiology*, vol. 217, pp. 251–256, 2000.

[37] J. Ko, H. Rusinek, E. Jacobs, J. Babb, M. Betke, G. McGuinness, and D. Naidich, "Small pulmonary nodules: Volume measurement at chest CT - phantom study," *Radiology*, vol. 228, pp. 864–870, 2003.

[38] C. Jaffe, "Measures of response: RECIST, WHO, and new alternatives," *J Clin Oncol*, vol. 24, pp. 3245–3251, 2006.

[39] M. Gavrielides, L. Kinnard, K. Myers, and N. Petrick, "Noncalcified lung nodules: Volumetric assessment with thoracic CT," *Radiology*, vol. 251, pp. 26–37, 2009.

[40] S. Armato III, F. Li, M. Giger, H. MacMahon, S. Sone, and K. Doi, "Lung cancer: performance of automated lung nodule detection applied to cancers missed in a CT screening program," *Radiology*, vol. 225, pp. 685–693, 2002.

[41] T. Messay, R. Hardie, and S. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," *Med Imag Anal*, vol. 14, pp. 390–406, 2010.

[42] Y. Kawata, N. Niki, H. Ohmatsu, R. Kakimuma, K. Eguchi, M. kaneko, and N. Moriyama, "Classification of pulmonary nodules in thin-section CT images based on shape characterization," in *Proc. IEEE Int. Conf. Image Processing*, 1997.

[43] K. Suzuki, F. Li, S. Sone, and K. Doi, "Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network," *IEEE Trans. Medical Imaging*, vol. 24, pp. 1138–1150, 2005.

[44] W. Kostis, D. Yankelevitz, A. Reeves, S. Fluture, and C. Henschke, "Small pulmonary nodules: reproducibility of three-dimensional volumetric measurement and estimation of time to follow-up CT," *Radiology*, vol. 231, pp. 446–452, 2004.

[45] D. Wormanns, G. Kohl, E. Klotz, A. Marheine, F. Beyer, W. Heindel, and S. Diederich, "Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility," *Chest*, vol. 14, pp. 86–92, 2004.

[46] J. Goo, T. Tongdee, R. Tongdee, K. Yeo, C. Hildebolt, and K. Bae, "Volumetric measurement of synthetic lung nodules with multi-detector row CT: Effect of various image reconstruction parameters and segmentation thresholds on measurement accuracy," *Radiology*, vol. 235, pp. 850–856, 2005.

[47] A. Reeves, A. Biancardi, T. Apanasovich, C. Meyer, H. MacMahon, E. van Beek, E. Kazerooni, D. Yankelevitz, M. McNitt-Gray, G. McLennan, S. Armato III, C. Henschke, D. Aberle, B. Croft, and L. Clarke, "The lung image database consortium (LIDC): A comparison of different size metrics for pulmonary nodule measurements," *Acad Radiol*, vol. 14, pp. 1475–1485, 2007.

[48] H. Ashraf, B. de Hoop, S. Shaker, A. Dirksen, K. Back, H. Hansen, M. Prokop, and J. Pedersen, "Lung nodule volumetry: segmentation algorithms within the same software package cannot be used interchangeably," *Euro Radiol*, 2010, in press.

[49] H. P.A., V. Romano, P. Rogalla, C. Klessen, A. Lembcke, L. Bomemann, V. Dicken, B. Hamm, and H. Bauknecht, "Variability of semiautomated lung nodule volumetry on ultralow-dose CT: Comparison with nodule volumetry on standard-dose CT," *J Digit Imag*, vol. 23, pp. 8–17, 2010.

[50] M. Rinaldi, T. Bartalena, L. Braccaioli, N. Sverzellati, S. Mattioli, E. Rimondi, G. Rossi, M. Zompatori, G. Battista, and R. Canini, "Three-dimensional analysis of pulmonary nodules: variability of semiautomated volume measurements between different versions of the same software," *Radiol med*, vol. 115, pp. 403–412, 2010.

[51] S. Sone, K. Tsushima, K. Yoshida, K. Hamanaka, T. Hanaoka, and R. Kondo, "Pulmonary nodules: Preliminary experience with semiautomated volumetric evaluation by CT stratum," *Acad Radiol*, 2010, in press.

[52] J. Bi, S. Periaswamy, K. Okada, T. Kubota, G. Fung, M. Salganicoff, and R. Rao, "Computer aided detection via asymmetric cascade of sparse hyperplane classifiers," in *ACM SIGKDD*, 2006, pp. 837–844.

[53] M. Lee, R. Wiemker, L. Boroczky, K. Sungur-Stasik, A. Cann, A. Borczuk, S. Kawut, and C. Powell, "Impact of segmentation uncertainties on computer-aided diagnosis of pulmonary nodules," *Int J CARS*, vol. 3, pp. 551–558, 2008.

[54] A. Witkin, "Scale-space filtering," in *Int. Joint. Conf. Artificial Intell.*, Karlsruhe, 1983, pp. 1019–1021.

[55] J. Koenderink, "The structure of images," *Biol. Cybern.*, vol. 50, pp. 363–370, 1984.

[56] L. Florack, B. Ter Haar Romey, J. Koenderink, and M. Viergever, "Cartesian differential invariants in scale-space," *J Math Imag Vis*, vol. 3, pp. 327–348, 1993.

[57] M. Nielsen, L. Florack, and R. Deriche, "Regularization, scale space, and edge detection filters," *J Math Imag Vis*, vol. 7, no. 4, pp. 291–307, 1997.

[58] J. Sporring, M. Nielsen, L. Florack, and P. Johansen, *Gaussian Scale-Space Theory*. Springer, 1997.

[59] A. Jirapatnakul, S. Fotin, A. Reeves, A. Biancardi, D. Yankelevitz, and C. I. Henschke, "Automated nodule location and size estimation using a multi-scale Laplacian of Gaussian filtering approach," in *Proc. IEEE Engineering in Medicine and Biology Society*, 2009.

[60] T. Lindeberg and J. Garding, "Shape-adapted smoothing in estimation of 3-D shape cues from affine distortions of local 2-D brightness structure," *Image and Vision Computing*, vol. 15, pp. 415–434, 1997.

[61] S. van Huffel and J. Vandewalle, *The Total Least Squares Problem Computational Aspects and Analysis*. Philadelphia: SIAM, 1991.

[62] Y. Chen and J. McInroy, "Estimating symmetric, positive definite matrices in robotic control," in *IEEE Int. Conf. Robotics and Automation*, Washington D.C., 2002, pp. 4269–4274.

[63] R. Malladi, J. Sethian, and B. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE Trans. Patt Anal Mach Intelli*, vol. 17, pp. 158–175, 1995.

[64] S. Kim, K. Yoon, and I. Kweon, "Object recognition using a generalized robust invariant feature and gestalt's law of proximity and similarity," vol. 41, pp. 726–741, 2008.

[65] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Patt Anal Mach Intelli*, vol. 12, no. 7, pp. 629–639, 1990.

[66] A. Almansa and T. Lindeberg, "Fingerprint enhancement by shape adaptation of scale-space operators with automatic scale selection," *IEEE Trans. Image Proc*, vol. 9, pp. 2027–2042, 2000.

[67] M. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Image Proc*, vol. 7, pp. 421–432, 1998.

[68] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Computer Vision*, vol. 60, pp. 91–110, 2004.

[69] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Computer Vision*, vol. 60, pp. 63–86, 2004.

[70] W. Freeman and E. Adelson, "The design and use of steerable filters," *IEEE Trans. Patt Anal Mach Intelli*, vol. 13, pp. 891–906, 1991.

[71] J. Bigun, G. Granlund, and J. Wiklund, "Multidimensional orientation estimation with applications to texture analysis and optical flow," *IEEE Trans. Patt Anal Mach Intelli*, vol. 13, pp. 775–790, 1991.

[72] F. Faas and L. van Vliet, "3D-orientation space; filters and sampling," in *Scandinavian Conf Image Anal*, 2003.

[73] C. Lampert and O. Wirjadi, "An optimal non-orthogonal separation of the anisotropic Gaussian convolution filter," *IEEE Trans. Image Proc*, vol. 15, pp. 3501–3513, 2006.

[74] A. Bhalerao and R. Wilson, "Estimating local and global structure using a Gaussian intensity model," in *Medical Image Understanding and Analysis*, 2001.

[75] R. Manmatha and N. Srimal, "Scale space technique for word segmentation in handwritten documents," in *Int Conf Scale-Space Theories in Computer Vision*, 1999.

[76] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J Roy Stats Soc Series B*, vol. 39, pp. 1–38, 1977.

[77] R. T. Collins, "Mean-shift blob tracking through scale space," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2003, pp. II:234–240.

[78] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Information Theory*, vol. 21, no. 1, pp. 32–40, 1975.

[79] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Patt Anal Mach Intelli*, vol. 17, no. 8, pp. 790–799, 1995.

[80] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2000, pp. 142–149.

[81] D. Comaniciu and P. Meer, "Mean shift analysis and applications," in *Proc. IEEE Int. Conf. Computer Vision*, 1999, pp. 1197–1203.

[82] M. P. Wand and M. C. Jones, *Kernel Smoothing*. Chapman & Hall, 1995.

[83] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Info. Theory*, vol. 37, no. 1, pp. 145–151, 1991.

[84] P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection*. New York: John Wiley, 1987.

[85] Y. Kanazawa and K. Kanatani, "Do we really have to consider covariance matrices for image features?" in *Proc. IEEE Int. Conf. Computer Vision*, Vancouver, 2001, pp. 586–591.