Scale Selection for Anisotropic Scale-Space: Application to Volumetric Tumor Characterization

Kazunori Okada, Dorin Comaniciu
Siemens Corporate Research, Inc.
Princeton, NJ, USA

Arun Krishnan
Siemens Medical Solutions USA, Inc.
Malvern, PA, USA

Abstract

A unified approach for treating the scale selection problem in the anisotropic scale-space is proposed. The anisotropic scale-space is a generalization of the classical isotropic Gaussian scale-space by considering the Gaussian kernel with a fully parameterized analysis scale (bandwidth) matrix. The "maximum-over-scales" and the "most-stable-over-scales" criteria are constructed by employing the "L-normalized scale-space derivatives", i.e., response-normalized derivatives in the anisotropic scale-space. This extension allows us to directly analyze the anisotropic (ellipsoidal) shape of local structures. The main conclusions are (i) the norm of the γ- and L-normalized anisotropic scale-space derivatives with a constant γ=1/2 are maximized regardless of the signal's dimension iff the analysis scale matrix is equal to the signal's covariance and (ii) the most-stable-over-scales criterion with the isotropic scale-space outperforms the maximum-over-scales criterion in the presence of noise. Experiments with 1D and 2D synthetic data confirm the above findings. 3D implementations of the most-stable-over-scales methods are applied to the problem of estimating anisotropic spreads of pulmonary tumors shown in high-resolution computed-tomography (HRCT) images. Comparison of the first- and second-order methods shows the advantage of exploiting the second-order information.

1. Introduction

Gaussian scale-space theory [12, 5, 2, 7, 9] offers a general paradigm for analyzing various local image features of arbitrary size. One of its useful attributes is the *maximum-over-scales property* of the γ-normalized derivatives [7], i.e., the characteristic scale of a feature located at the local spatial maximum corresponds to the bandwidth of the Gaussian kernel that provides the local maximum of the normalized derivatives over the varying bandwidths. This is a solution to the general scale selection problem: *given a set of analysis scales (bandwidths), find the analysis scale that provides the best estimate of the local feature's scale or other properties*. The theory has been studied extensively and applied to various problems. However, the main focuses have been on the scale-space functions which model either the isotropic homogeneous [12, 5] or anisotropic inhomogeneous [11] diffusion processes.

This paper introduces *anisotropic scale-space* as a solution to the *anisotropic homogeneous* diffusion equation. The anisotropic scale-space is characterized by a fully parameterized analysis scale matrix and is a generalization of the classical isotropic Gaussian scale-space. This extension allows us to directly analyze the anisotropic (ellipsoidal) shape of the local structures, i.e., *the scale-space analysis can be interpreted as the covariance estimation of signals locally modeled by a Gaussian-based function*. Such an anisotropic scale-space construction has scarcely been studied in the context of the scale selection problem.

We propose two novel scale selection frameworks, *maximum-over-scales* and *most-stable-over-scales* criteria, constructed from *L-normalized scale-space derivatives* that are response-normalized derivatives in the anisotropic scale-space. For practical consideration, the isotropic scale-space is employed for constructing the most-stable-over-scales criterion. By considering local Gaussian-like (blob-like) structures, we derive a number of scale selection solutions from the first- and second-order normalized derivatives. In many applications, the second-order blob feature provides essential information of target structures such as tumors in medical imaging or faces in surveillance applications. Its usefulness has also been discussed in a recent study by Lillholm et al. [6].

The main conclusions of this paper are (i) the norm of the γ- and L-normalized anisotropic scale-space derivatives with a constant γ=1/2 exhibit the maximum-over-scales property regardless of the signal's dimension for both the use of the first- and second-order derivatives and (ii) the most-stable-over-scales criterion with the isotropic scale-space outperforms the maximum-over-scales criterion in the presence of noise. Additionally, we discuss the relationship and equivalence of some of the proposed methods to the previously proposed covariance estimation methods [4, 1]. Experiments with 1D and 2D synthetic data are conducted

to validate these findings. Finally, we apply 3D implementations of the proposed methods to the problem of estimating anisotropic spreads of pulmonary tumors shown in high-resolution computed-tomography (HRCT) images. Comparison of the first- and second-order methods indicates the advantage of exploiting the second-order information. The results of these experiments are given in Sec. 5.

2. Anisotropic Scale-Space

This section provides definitions of the anisotropic scale-space and its derivatives. Given a d-variate continuous positive signal $f(\mathbf{x})$, the local region of f forming a Gaussian-like structure around a spatial local maximum $\mathbf{u}$ can be approximated by a product of a d-variate Gaussian function and a positive multiplicative parameter,

$$f(\mathbf{x}) \simeq \alpha \times \Phi(\mathbf{x}; \mathbf{u}, \mathbf{\Sigma})|_{\mathbf{x}\in\mathcal{S}} \tag{1}$$

where $\mathcal{S}$ is a set of data points in the neighborhood of $\mathbf{u}$, belonging to the basin of attraction of $\mathbf{u}$, and $\Phi(\mathbf{x}; \mathbf{u}, \mathbf{\Sigma}) = (2\pi)^{-d/2}|\mathbf{\Sigma}|^{-1/2}\exp(-\frac{1}{2}(\mathbf{x}-\mathbf{u})^t\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{u}))$. The covariance $\mathbf{\Sigma}$ of Φ describes the spread of the local Gaussian-like structure. Its anisotropy can be specified only by a fully parameterized covariance.

The Gaussian scale-space is a one-parameter family of a d-variate continuous signal f provided by a convolution with isotropic Gaussian kernels $\Phi(\mathbf{x}; \mathbf{0}, \mathbf{H} = h\boldsymbol{I})$ with varying analysis scales (bandwidths) $h \geq 0$.

$$L(\mathbf{x}; \mathbf{H}) \equiv f(\mathbf{x}) * \Phi(\mathbf{x}; \mathbf{0}, \mathbf{H}) \tag{2}$$

Such a linear scale-space is known to be a solution of the isotropic diffusion equation $\partial_h L = 1/2\nabla^2 L$ [12, 5].

We define the *anisotropic scale-space* as a generalization of Eq.(2) by considering a fully parameterized symmetric positive definite *analysis scale matrix* $\mathbf{H} \in \mathcal{SPD} \in \mathcal{R}^{d\times d}$, where $\mathcal{SPD}$ denotes the set of symmetric positive definite matrices. The anisotropic scale-space is a solution to the anisotropic homogeneous diffusion [1] $\partial_{\mathbf{H}} L = 1/2\nabla\nabla^t L$.

2.1. Scale-Space Derivatives

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 scale-space gradient vector $G(\mathbf{x}; \mathbf{H}) \in \mathcal{R}^d$ and scale-space Hessian matrix $\mathbf{P}(\mathbf{x}; \mathbf{H}) \in \mathcal{R}^{d\times d}$ are defined by,

$$\begin{aligned} G(\mathbf{x}; \mathbf{H}) &\equiv \nabla L(\mathbf{x}; \mathbf{H}) \\ &= f(\mathbf{x}) * \Phi(\mathbf{x}; \mathbf{H})\mathbf{H}^{-1}(-\mathbf{x}) \qquad (3) \\ \mathbf{P}(\mathbf{x}; \mathbf{H}) &\equiv \nabla\nabla^t L(\mathbf{x}; \mathbf{H}) \\ &= f(\mathbf{x}) * \Phi(\mathbf{x}; \mathbf{H})\mathbf{H}^{-1}(\mathbf{x}\mathbf{x}^t - \mathbf{H})\mathbf{H}^{-1} \qquad (4) \end{aligned}$$

[1] It should not be confused with the well-known anisotropic diffusion [11] which models inhomogeneous diffusion processes.

On the other hand, by substituting Eq.(1) to Eq.(2), Eq.(3), and Eq.(4), analytical formula of the scale-space L and its derivatives G and $\mathbf{P}$ are derived as functions of a Gaussian with a covariance matrix $\mathbf{\Sigma} + \mathbf{H}$,

$$\begin{aligned} L(\mathbf{x}; \mathbf{H}) &= \alpha\Phi(\mathbf{x}; \mathbf{u}, \mathbf{\Sigma} + \mathbf{H}) \qquad (5) \\ G(\mathbf{x}; \mathbf{H}) &= \alpha\Phi(\mathbf{x}; \mathbf{u}, \mathbf{\Sigma} + \mathbf{H})(\mathbf{\Sigma} + \mathbf{H})^{-1}(\mathbf{u} - \mathbf{x}) \qquad (6) \\ \mathbf{P}(\mathbf{x}; \mathbf{H}) &= \alpha\Phi(\mathbf{x}; \mathbf{u}, \mathbf{\Sigma} + \mathbf{H})(\mathbf{\Sigma} + \mathbf{H})^{-1} \qquad (7) \\ &\quad [(\mathbf{u} - \mathbf{x})(\mathbf{u} - \mathbf{x})^t - (\mathbf{\Sigma} + \mathbf{H})](\mathbf{\Sigma} + \mathbf{H})^{-1} \end{aligned}$$

2.2. L-normalized Scale-Space Derivatives

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

$$G_l(\mathbf{x}; \mathbf{H}) \equiv \frac{G(\mathbf{x}; \mathbf{H})}{L(\mathbf{x}; \mathbf{H})} = (\mathbf{\Sigma} + \mathbf{H})^{-1}(\mathbf{u} - \mathbf{x}) \tag{8}$$

$$\begin{aligned} \mathbf{P}_l(\mathbf{x}; \mathbf{H}) &\equiv \frac{\mathbf{P}(\mathbf{x}; \mathbf{H})}{L(\mathbf{x}; \mathbf{H})} \qquad (9) \\ &= (\mathbf{\Sigma} + \mathbf{H})^{-1}(\mathbf{u} - \mathbf{x})(\mathbf{u} - \mathbf{x})^t(\mathbf{\Sigma} + \mathbf{H})^{-1} \\ &\quad -(\mathbf{\Sigma} + \mathbf{H})^{-1} \end{aligned}$$

They are response-normalized derivatives in the scale-space and vanish both the multiplicative parameter and the exponential term from the derivative formulae. Both L-normalized scale-space gradient and Hessian are computable since $L(\mathbf{x}; \mathbf{H})$ is non-zero within a finite range with positive $f(\mathbf{x})$.

3. Maximum-Over-Scales Criterion

The maximum-over-scales criterion was proposed by Lindeberg [7] by using the γ-normalized derivatives. Given a Gaussian scale-space, some scale-space derivative functions normalized by the analysis scale raised to the power of an upper-bounded real value γ assume their local maximum at the characteristic scale of the target feature. For the d-variate local Gaussian-like structures, the γ-normalized Laplacian with $\gamma = (d+2)/4$ evaluated at a spatial local maximum, $\mathrm{tr}(\mathbf{H}^{(d+2)/4}\mathbf{P}(\mathbf{u}; \mathbf{H}))$, is locally maximized over scales when the analysis scale h is equal to the signal's variance σ^2, where "tr" denotes the trace of a $d \times d$ matrix, $\mathbf{H} = h\boldsymbol{I}$, and $\mathbf{\Sigma} = \sigma^2\boldsymbol{I}$. Our pilot study showed that this maximum-over-scales property holds for the anisotropic scale-space with fully parameterized $\mathbf{H}$ and $\mathbf{\Sigma}$.

We develop a new maximum-over-scales criterion constructed with the norm of the γ-normalization of the L-normalized scale-space derivatives. Note that γ in the Lindeberg's criterion depends on the dimension of the signal

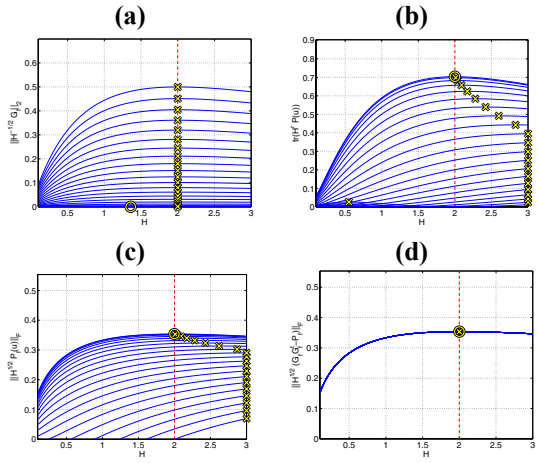

Figure 1: Examples of the maximum-over-scales methods. A centered 1D Gaussian signal with $\sigma^2 = 2$ is used as target. (a) the first-order method with Eq.(10), (b) the γ-normalized Laplacian, (c) the second-order method with Eq.(11), (d) the second-order method with Eq.(12). Curved lines in each 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.

for assuming the maximum-over-scales property. The proposed criteria provide an elegant solution, in which a constant γ = 1/2 gives rise to the maximum-over-scales property regardless of the signal's dimensions for both the first- and second-order cases. The spatial local maximum location $\mathbf{u}$ is assumed to be known hereafter. For notational simplicity, the function arguments of $G_l(\mathbf{x};\mathbf{H})$ and $\mathbf{P}_l(\mathbf{x};\mathbf{H})$ are omitted unless they are evaluated at a specific location. Fig.1 illustrates the proposed criteria with a synthetic 1D Gaussian signal.

3.1. First-Order Method

Using Eq.(8), a γ-normalization of the L-normalized scale-space gradient vector with γ=1/2 is expressed by $\mathbf{H}^{1/2}G_l$ = $\mathbf{H}^{1/2}(\mathbf{\Sigma}+\mathbf{H})^{-1}(\mathbf{u}-\mathbf{x})$. We consider L_2 norm of the normalized gradient,

$$\|\mathbf{H}^{1/2}G_l\|_2 = \|\mathbf{H}^{1/2}(\mathbf{\Sigma}+\mathbf{H})^{-1}(\mathbf{u}-\mathbf{x})\|_2 \qquad (10)$$

Rewriting this equation with the mean shift vector $\mathbf{m}(\mathbf{x};\mathbf{H})$ $\equiv \mathbf{H}G_l$ [1, 10] results in $\|\mathbf{H}^{1/2}G_l\|_2 = \|\mathbf{H}^{-1/2}\mathbf{m}\|_2$. This demonstrates that the L_2 norm is equivalent with the *magnitude of the bandwidth-normalized mean shift vector* introduced in Theorem 1 of [1, p.282]. The theorem states that such magnitude exhibits the maximum-over-scales property with $\mathbf{H} = \mathbf{\Sigma}$. The proof is provided in [1, p.287]. Thus the L_2 norm of the γ- and L-normalized scale-space gradient vector possesses the maximum-over-scales property. This criterion holds at arbitrary locations $\mathbf{x} \in \mathcal{S}$ except at $\mathbf{u}$ as shown in Fig.1a.

3.2. Second-Order Methods

Two types of second-order scale selection methods are considered. First, a solution only with the Hessian matrix is examined. Using Eq.(9), the γ-normalization of the L-normalized scale-space Hessian matrix with γ=1/2 is expressed by $\mathbf{H}^{1/2}\mathbf{P}_l = \mathbf{H}^{1/2}G_lG_l^t - \mathbf{H}^{1/2}(\mathbf{\Sigma}+\mathbf{H})^{-1}$. When evaluated at the spatial maximum $\mathbf{u}$, the normalized Hessian is reduced to the following form since G_l becomes zero: $\mathbf{H}^{1/2}\mathbf{P}_l(\mathbf{u};\mathbf{H}) = -\mathbf{H}^{1/2}(\mathbf{\Sigma}+\mathbf{H})^{-1}$. The Frobenius matrix norm of this derivative matrix function is given by,

$$\|\mathbf{H}^{1/2}\mathbf{P}_l(\mathbf{u};\mathbf{H})\|_F \quad = \quad \|\mathbf{H}^{1/2}(\mathbf{\Sigma}+\mathbf{H})^{-1}\|_F \quad (11)$$

The following maximum-over-scales method is obtained using Eq.(11),

Proposition 1 The Frobenius norm of the γ- and L-normalized scale-space Hessian matrix with γ = 1/2 is maximized when the fully parameterized analysis scale matrix $\mathbf{H} \in \mathcal{SPD}$ is equal to $\mathbf{\Sigma}$.

Proof given in Appendix A.

Proposition 1 is true only at the spatial maximum $\mathbf{u}$ as shown in Fig.1c. The d-variate γ-normalized Laplacian can be expressed as a matrix trace: $\mathrm{tr}(\mathbf{H}^{(d+2)/4}\mathbf{P}(\mathbf{u};\mathbf{H}))$ = $-L(\mathbf{u};\mathbf{H})\mathrm{tr}(\mathbf{H}^{(d+2)/4}(\mathbf{\Sigma}+\mathbf{H})^{-1})$. Also the Frobenius norm in Eq.(11) can be expressed by $\|\mathbf{H}^{1/2}\mathbf{P}_l(\mathbf{u};\mathbf{H})\|_F^2$ = $\mathrm{tr}((\mathbf{\Sigma}+\mathbf{H})^{-1}\mathbf{H}(\mathbf{\Sigma}+\mathbf{H})^{-1})$. As compared in Fig.1b and Fig.1c, both methods behave similarly despite the difference in their functional forms.

Second, a solution that includes both gradient and Hessian is examined. From Eq.(9), the γ- and L-normalization of a derivative function $G_lG_l^t - \mathbf{P}_l$ with γ=1/2 is given by $\mathbf{H}^{1/2}(G_lG_l^t - \mathbf{P}_l)$ = $\mathbf{H}^{1/2}(\mathbf{\Sigma}+\mathbf{H})^{-1}$. And its Frobenius norm is,

$$\|\mathbf{H}^{1/2}(G_lG_l^t - \mathbf{P}_l)\|_F = \|\mathbf{H}^{1/2}(\mathbf{\Sigma}+\mathbf{H})^{-1}\|_F \qquad (12)$$

Consequently, we obtain the following maximum-over-scales method,

Proposition 2 Consider a scale-space derivative matrix function of a sum of the outer-product of the L-normalized scale-space gradient vector and the negated L-normalized scale-space Hessian matrix. The Frobenius norm of the γ-normalization of this matrix function with γ = 1/2 is maximized when the fully parameterized analysis scale matrix $\mathbf{H} \in \mathcal{SPD}$ is equal to $\mathbf{\Sigma}$.

Proof given in Appendix B.

As shown in Fig.1d, this solution is *invariant* against the locations and its maximum-over-scales property holds for all the locations $\mathbf{x} \in \mathcal{S}$.

4. Most-Stable-Over-Scales Criterion

We develop the most-stable-over-scales criterion constructed again by employing the L-normalized scale-space derivatives. This approach exploits the fact that the scale selection with the anisotropic scale-space can be seen as fully parameterized covariance estimation. Each derived method consists of two steps: 1) least-squares estimation of the signal's covariance $\mathbf{\Sigma}(h)$ for each isotropic analysis scale h and 2) divergence-based stability test for obtaining the most stable estimate over the scales $\hat{\mathbf{\Sigma}} = \mathbf{\Sigma}(h^* = \text{argmin div}\{\mathbf{\Sigma}(h)\})$.

The maximum-over-scales criterion 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. For this practical reason, this criterion employs *isotropic* analysis scales $\mathbf{H} = h\boldsymbol{I}$ $(h \in \mathcal{R} > 0)$. This is possible because the direct covariance estimators described in the next section are valid with arbitrary scale matrices $\mathbf{H}$. Furthermore, estimation errors due to noise can be reduced by combining a set of estimates derived from different locations within the basin of attraction of $\mathbf{u}$ since the direct estimators are also satisfied at arbitrary locations $\mathbf{x} \in \mathcal{S}$.

For the stability test, a form of the Jensen-Shannon divergence proposed in [1] is employed given a set of ordered analysis scales forming a geometric sequence $\{h_s | s = 1, .., S; \frac{h_{s+1}}{h_s} = constant\}$,

$$\begin{aligned} JS(s) &= \frac{1}{2}\log\frac{|\frac{1}{2a+1}\sum_{s-a}^{s+a}\mathbf{\Sigma}(h_i)|}{\sqrt[2a+1]{\prod_{s-a}^{s+a}|\mathbf{\Sigma}(h_i)|}} \\ &+ \frac{1}{2}\sum_{s-a}^{s+a}(\mathbf{u}(h_i)-\mathbf{u})^t(\sum_{s-a}^{s+a}\mathbf{\Sigma}(h_i))^{-1}(\mathbf{u}(h_i)-\mathbf{u}) \end{aligned} \quad (13)$$

where $\mathbf{u} = \frac{1}{2a+1}\sum_{s-a}^{s+a}\mathbf{u}(h_i)$ and a is a neighborhood parameter.

4.1. Direct Covariance Estimators

This section derives explicit estimators of the signal's covariance $\mathbf{\Sigma}$ used for constructing the stability-based scale selection criterion. The L-normalized scale-space derivatives can be numerically computed from the given signal $f(\mathbf{x})$ by using Eq.(2), Eq.(3), and Eq.(4). The resulting equations are satisfied with any given anisotropic analysis scale matrices $\mathbf{H} \in \mathcal{SPD}$.

A covariance estimator with the normalized gradient G_l is derived by manipulating Eq.(8) while maintaining its equality,

$$\mathbf{\Sigma}G_l = \mathbf{u} - \mathbf{x} - \mathbf{H}G_l \quad (14)$$

The resulting equation of an unknown $\mathbf{\Sigma}$ is under-complete, requiring at least two independent samples for the unique solution. Given a sufficient number of independent samples, an over-complete normal equation can be formed and solved by a constrained least-squares method. 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 [1, 10], i.e., $\mathbf{\Sigma}\mathbf{H}^{-1}\mathbf{m} = \mathbf{u}-\mathbf{x}-\mathbf{m}$. This assumes exactly the same form as the constrained least-squares formulation proposed in [10]. Both equations become singular when G_l goes to zero at $\mathbf{x} = \mathbf{u}$.

Another covariance estimator with the normalized Hessian $\mathbf{P}_l$ is derived by manipulating Eq.(9) while maintaining its equality,

$$\mathbf{\Sigma} = (G_l G_l^t - \mathbf{P}_l)^{-1} - \mathbf{H} \quad (15)$$

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.(15) collapses into the form only with the Hessian matrix,

$$\begin{aligned} \mathbf{\Sigma} &= (-\mathbf{P}_l(\mathbf{u};\mathbf{H}))^{-1} - \mathbf{H} \\ &= L(\mathbf{u};\mathbf{H})(-\mathbf{P}(\mathbf{u};\mathbf{H}))^{-1} - \mathbf{H} \end{aligned} \quad (16)$$

The resulting form is similar to the well-known Hessian-based covariance estimator [4, 8], 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: $\mathbf{P}(\mathbf{u};\mathbf{H}) = -\alpha(2\pi)^{-d/2}|\mathbf{\Sigma}+\mathbf{H}|^{-1/2}(\mathbf{\Sigma}+\mathbf{H})^{-1}$. This equation can be solved for $\mathbf{\Sigma}$ since $\mathbf{H}+\mathbf{\Sigma} \in \mathcal{SPD}$, i.e., $\mathbf{\Sigma} = \alpha^{\frac{2}{d+2}}|2\pi(-\mathbf{P}(\mathbf{u};\mathbf{H}))^{-1}|^{-\frac{1}{d+2}}(-\mathbf{P}(\mathbf{u};\mathbf{H}))^{-1} - \mathbf{H}$. Since this and Eq.(16) must be equivalent, the following formula is obtained after some algebra,

$$\alpha = \sqrt{|2\pi(-\mathbf{P}(\mathbf{u};\mathbf{H}))^{-1}|L(\mathbf{u};\mathbf{H})^{d+2}} \quad (17)$$

The scale-space Hessian $\mathbf{P}(\mathbf{x};\mathbf{H})$ is symmetric *negative* definite if $\mathbf{x}$ is at a stable critical point of $-L(\mathbf{x};\mathbf{H})$. When $\mathbf{P}(\mathbf{u};\mathbf{H})$ is numerically computed by using Eq.(4), it must be assured that $\mathbf{u}$ satisfies this condition so that the estimated $\mathbf{\Sigma}$ by Eq.(16) satisfies the positive definite constraint and Eq.(17) remains as real-valued.

4.2. First-Order Method

The first-order most-stable-over-scales method exploiting the direct covariance estimator of Eq.(14) takes the same procedure as our previously proposed mean shift-based solution [10], Following briefly describes the method.

Given the spatial maximum locations $\mathbf{u}(h)$ in $L(\mathbf{x}; h\boldsymbol{I})$, we sample a set of K measurement pairs $\{(\mathbf{x}_k, G_l(\mathbf{x}_k; h\boldsymbol{I}))|k = 1, .., K\}$ within

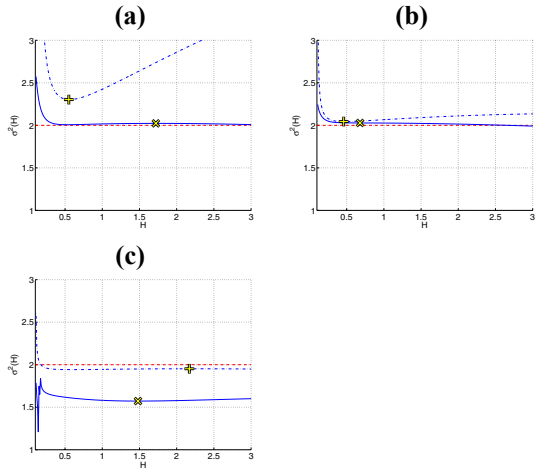

Figure 2: Examples of the most-stable-over-scales methods. A centered 1D Gaussian signal of $\sigma^2 = 2$ with additive random noise ($randn * 0.01$) is used as target. At each analysis scale, the variance of the target is estimated from a set of samples within: (a) $\pm 0.1\sigma$, (b) $\pm 1.0\sigma$, and (c) $\pm 3.0\sigma$. Dash lines: the ground-truth scale. Dot lines: the first-order estimates by Eq.(14). Solid lines: the second-order estimates by Eq.(15). "+" and "×" denote the most stable estimates by the first- and second-order methods, respectively.

the basin of attraction of $\mathbf{u}$. These samples are used to construct an over-complete normal equation $\mathbf{A}\boldsymbol{\Sigma} = \mathbf{B}$ where $\mathbf{A} \equiv (G_{f1}, .., G_{fK})^t$ and $\mathbf{B} \equiv (\mathbf{u} - \mathbf{x}_1 - hG_{f1}, .., \mathbf{u} - \mathbf{x}_K - hG_{fK})^t$. The constrained least-squares solution of the normal equation for the unknown $\boldsymbol{\Sigma} \in \mathcal{SPD}$ is given by finding the minimizer $\mathbf{Y}^*$ of an area criterion $\|\mathbf{AY} - \mathbf{BY}^{-t}\|_F^2$ where $\mathbf{Y}$ is Cholesky factorization of $\boldsymbol{\Sigma} = \mathbf{YY}^t$. The closed-form of this solution is expressed by a function of symmetric Schur decompositions of $\mathbf{P} \equiv \mathbf{A}^t\mathbf{A}$ and $\tilde{\mathbf{Q}} \equiv \boldsymbol{\Sigma}_P\mathbf{U}_P^t\mathbf{Q}\mathbf{U}_P\boldsymbol{\Sigma}_P$ given $\mathbf{Q} \equiv \mathbf{B}^t\mathbf{B}$,

$$\begin{aligned}\boldsymbol{\Sigma} &= \mathbf{U}_P\boldsymbol{\Sigma}_P^{-1}\mathbf{U}_{\tilde{Q}}\boldsymbol{\Sigma}_{\tilde{Q}}\mathbf{U}_{\tilde{Q}}^t\boldsymbol{\Sigma}_P^{-1}\mathbf{U}_P^t \\ \mathbf{P} &= \mathbf{U}_P\boldsymbol{\Sigma}_P^2\mathbf{U}_P^t \\ \tilde{\mathbf{Q}} &= \mathbf{U}_{\tilde{Q}}\boldsymbol{\Sigma}_{\tilde{Q}}^2\mathbf{U}_{\tilde{Q}}^t\end{aligned} \tag{18}$$

Applying these equations to a given set of analysis scales results in a set of estimates $\{(\mathbf{u}(h), \boldsymbol{\Sigma}(h))\}$. The most stable estimate is found by the stability test with the Jensen-Shannon divergence in Eq.(13).

4.3. Second-Order Method

The second-order most-stable-over-scales method exploits the direct estimator of Eq.(15) or Eq.(16). Similar to the first-order method, we sample a set of K measurement pairs $\{(G_l(\mathbf{x}_k; h\boldsymbol{I}), \mathbf{P}_l(\mathbf{x}_k; h\boldsymbol{I}))\}$ within the neighborhood of $\mathbf{u}$. A least-squares covariance estimator is given by averaging

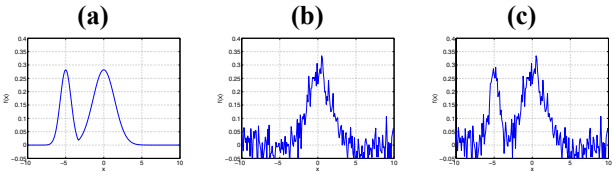

Figure 3: 1D synthetic data with noise. The target is the centered 1D Gaussian with $\sigma^2 = 2$. (a) a Gaussian centered at $\mathbf{u} = -5$ with $\sigma^2 = 0.5$ superimposed to the target. (b) the target with additive random noise ($randn * 0.04$). (c) the data (a) with the same additive noise in (b).

the local estimates,

$$\boldsymbol{\Sigma} = \frac{1}{K}\sum_{k=1}^{K}\{(G_l(\mathbf{x}_k; h\boldsymbol{I})G_l(\mathbf{x}_k; h\boldsymbol{I})^t - \mathbf{P}_l(\mathbf{x}_k; h\boldsymbol{I}))^{-1} - h\boldsymbol{I}\} \tag{19}$$

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.(16). The stability-based scale selection is achieved by the same manner as the first-order method,

$$\begin{aligned}\hat{\mathbf{u}} &= \mathbf{u}(h^*) \\ \hat{\boldsymbol{\Sigma}} &= \boldsymbol{\Sigma}(h^*) \\ h^* &= \text{argmin } JS(\mathbf{u}(h), \boldsymbol{\Sigma}(h))\end{aligned} \tag{20}$$

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 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.

5. Experiments

5.1. 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 method is more sensitive to the random noise, (iii) the second-order methods are more sensitive to

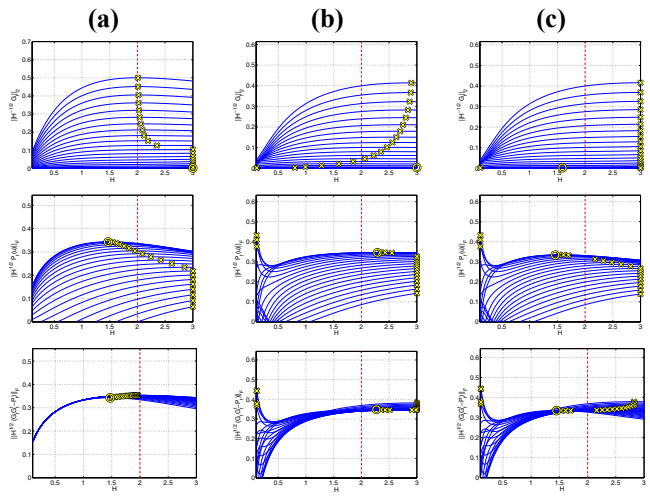

Figure 4: Variance estimation by the maximum-over-scales criterion for the signals shown in Fig.3. 1st row: the first-order method with Eq.(10). 2nd row: the second-order method with Eq.(11). 3rd row: the second-order method with Eq.(12). The legend is the same as Fig.1.

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 (dot lines) and second-order (solid lines) methods are compared by using the same data as Fig.4. 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 "$\times$" 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.

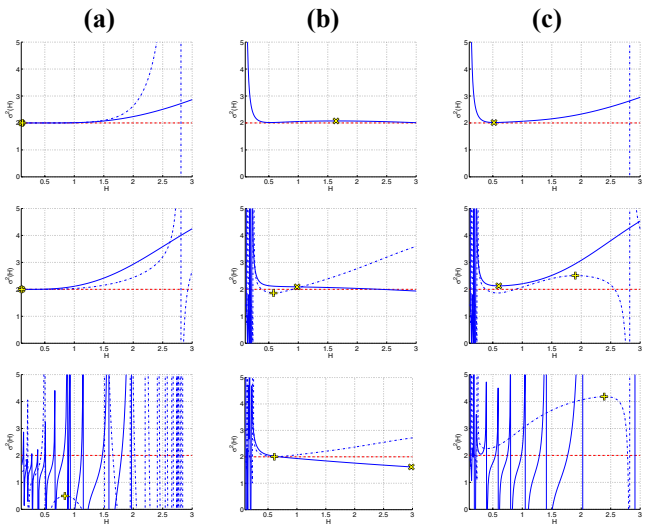

Figure 5: Variance estimation by the most-stable-over-scales methods for the signals in Fig.3. At each analysis scale, the variance of the target is estimated from samples within: 1st row: $\pm 0.1\sigma$, 2nd row: $\pm 1.0\sigma$, 3rd row: $\pm 2.0\sigma$. The legend is the same as Fig.2.

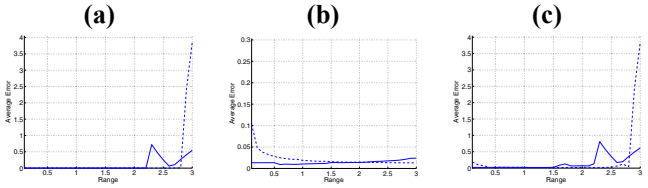

Figure 6: Average variance estimation errors of the most-stable-over-scales methods over 100 independent tests. The same types of data in Fig.3 with different random noise are used for each test. The errors are plotted against varying sampling ranges. Dot and solid lines denote errors by the first- and the second-order methods, respectively.

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.7a. Fig.7b-e show results with the maximum-over-scales methods. We use a set of 144 analysis scale matrices sampled along the two eigenvectors of the ground-truth matrix by setting the correspond-

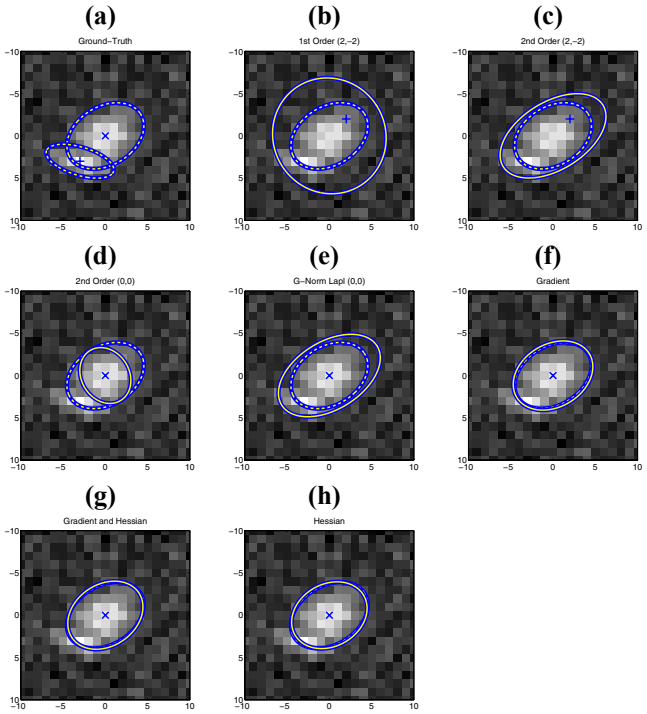

Figure 7: Examples with 2D synthetic data consisting of a target centered Gaussian with a neighboring structure centered at (-3,3) and additive random noise as shown in (a). From (b) to (e): the maximum-over-scales methods. (b) first-order method evaluated at (2,-2), (c) second-order method at the non-maximum 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.

ing eigenvalues to $(1, 2, .., 12)$. Fig.7f-h show results with the most-stable-over-scales methods. We use 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).

5.2. Lung CT Data

3D implementations of the most-stable-over-scales methods are applied to the problem of estimating anisotropic spreads

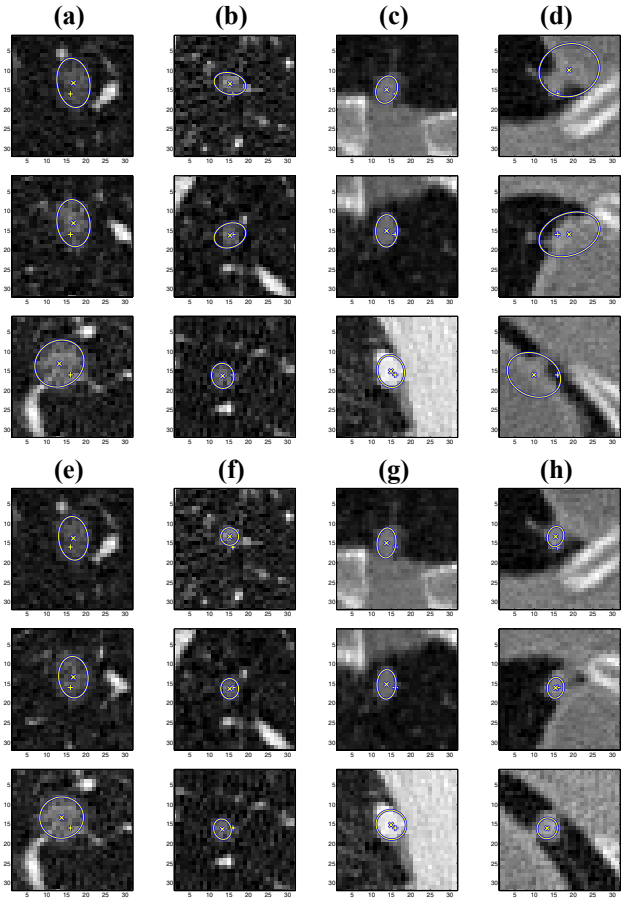

Figure 8: Examples of the 3D spread estimation of lung tumors in 3D HRCT data by using the most-stable-over-scales scale selection methods. "+" denotes the marker locations. The local spatial maxima and 3D spread estimates are denoted by "$\times$" and 2D intersections of 50% confidence ellipsoids, respectively. Cases (a)-(d) show the results by the first-order method. Cases (e)-(h) show those by the second-order (Hessian only) method.

of pulmonary tumors shown in high-resolution computed-tomography (HRCT) images. Each volumetric image consists of 12-bit positive values over an array of 512$\times$512 lattices. 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 locations of the local spatial maxima $\mathbf{u}(h)$ are estimated by using the mean shift-based mode seeking algorithm with the extended mean shift vector [1, 10]. Markers indicating rough tumor locations are given *a priori*. The convergence point of the majority of data points sampled around the marker provides the spatial maximum estimate $\mathbf{u}(h)$. The neighborhood width of the divergence formula is set to $a = 1$. 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.

HRCT data of 14 patients displaying the total of 77 pulmonary tumors were used for this evaluation. 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 [3]. The right columns show the on-the-wall cases. 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)).

6. Conclusions

We propose a unified approach for treating the scale selection problem in the anisotropic scale-space evaluating the local Gaussian-like structures, resulting in a number of the first- and second-order solutions. The maximum-over-scales criterion with the L-normalized anisotropic scale-space derivatives offers elegant scale selection solutions with the constant γ value, exploiting the analytical simplicity of the Gaussian function. For realistic application scenarios with the presence of noise, our experimental results demonstrate that the second-order most-stable-over-scales methods with the isotropic scale-space outperform others. For our future work, we plan to consider different types of features and formally analyze the proposed criteria with the arbitrary order of the L-normalized derivatives. The 3D tumor spread analysis system developed in this work provides the estimation of the tumor volumes and contours. We also plan to further improve the estimation accuracy by considering domain-specific hybrid models.

Acknowledgments

The authors wish to thank Visvanathan Ramesh from Siemens Corporate Research for stimulating discussions, Alok Gupta from CAD group, Siemens Medical Solutions, for his support and encouragement, and Jonathan Stoeckel from CAD group, Siemens Medical Solutions, for his valuable technical supports.

A. Proof of Proposition 1

We define $\eta(\mathbf{H}) \equiv \|\mathbf{H}^{1/2}\mathbf{P}_l(\mathbf{u};\mathbf{H})\|_F$. The proposition 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,

$$\begin{aligned}
&\eta(\mathbf{\Sigma})^2 - \eta(\mathbf{H})^2 \\
&= \|\mathbf{\Sigma}^{1/2}(\mathbf{\Sigma}+\mathbf{\Sigma})^{-1}\|_F^2 - \|\mathbf{H}^{1/2}(\mathbf{\Sigma}+\mathbf{H})^{-1}\|_F^2 \\
&= \tfrac{1}{4}\mathrm{tr}(\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2}) - \mathrm{tr}((\mathbf{\Sigma}+\mathbf{H})^{-1}\mathbf{H}(\mathbf{\Sigma}+\mathbf{H})^{-1}) \\
&= \tfrac{1}{4}\mathrm{tr}(\mathbf{\Sigma}^{-1} - 4(\mathbf{\Sigma}+\mathbf{H})^{-1}\mathbf{H}(\mathbf{\Sigma}+\mathbf{H})^{-1}) \\
&= \tfrac{1}{4}\mathrm{tr}((\mathbf{\Sigma}+\mathbf{H})^{-1}(\mathbf{H}\mathbf{\Sigma}^{-1} - \boldsymbol{I})^2\mathbf{\Sigma}(\mathbf{\Sigma}+\mathbf{H})^{-1})
\end{aligned} \tag{21}$$

Since $\mathbf{\Sigma}$ and $\mathbf{H}$ are positive definite, all the matrices inside the trace in Eq.(21) 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 $\mathbf{H} = \mathbf{\Sigma}$. □

B. Proof of Proposition 2

For all $\mathbf{x} \in \mathcal{S}$, $\|\mathbf{H}^{1/2}(G_l G_l^t - \mathbf{P}_l)\|_F = \|\mathbf{H}^{1/2}(\mathbf{\Sigma} + \mathbf{H})^{-1}\|_F = \eta(\mathbf{H})$. From the proof of Proposition 1, we have $\eta(\mathbf{\Sigma})^2 - \eta(\mathbf{H})^2 \geq 0$ with equality $\mathbf{H} = \mathbf{\Sigma}$. □

References

[1] D. Comaniciu. An algorithm for data-driven bandwidth selection. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(2):281–288, 2003.

[2] L. Florack, B. Ter Haar Romey, J. Koenderink, and M. Viergever. Cartesian differential invariants in scale-space. *J. Math. Imaging and Vision*, 3:327–348, 1993.

[3] 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.*, 178(5):1053–1057, 2002.

[4] Y. Kanazawa and K. Kanatani. Do we really have to consider covariance matrices for image features? In *Int. Conf. Computer Vision*, pages 586–591, Vancouver, 2001.

[5] J. Koenderink. The structure of images. *Biol. Cybern.*, 50:363–370, 1984.

[6] M. Lillholm, M. Nielsen, and L. Griffin. Feature-based image analysis. *Int. J. Comput. Vision*, 52(2/3):73–95, 2003.

[7] T. Lindeberg. Feature detection with automatic scale selection. *Int. J. Comput. Vision*, 30(2):79–116, 1998.

[8] B. Matei. *Heteroscedastic Errors-In-Variables Models in Computer Vision*. PhD thesis, Rutgers University, 2001.

[9] M. Nielsen, L. Florack, and R. Deriche. Regularization, scale space, and edge detection filters. *J. Mathematical Imaging and Vision*, 7(4):291–307, 1997.

[10] K. Okada, D. Comaniciu, N. Dalal, and A. Krishnan. A robust algorithm for characterizing anisotropic local structures. In *Eur. Conf. Computer Vision*, Prague, 2004.

[11] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(7):629–639, 1990.

[12] A. Witkin. Scale-space filtering. In *Int. Joint. Conf. Artificial Intell.*, pages 1019–1021, Karlsruhe, 1983.