

A Theory of Phase Singularities for Image Representation and its Applications to Object Tracking and Image Matching

Yu Qiao, *Member, IEEE*, Wei Wang, Nobuaki Minematsu, *Member, IEEE*, Jianzhuang Liu, *Senior Member, IEEE*, Mitsuo Takeda, and Xiaoou Tang, *Fellow, IEEE*

Abstract—This paper studies phase singularities (PSs) for image representation. We show that PSs calculated with Laguerre-Gauss filters contain important information and provide a useful tool for image analysis. PSs are invariant to image translation and rotation. We introduce several invariant features to characterize the core structures around PSs and analyze the stability of PSs to noise addition and scale change. We also study the characteristics of PSs in a scale space, which lead to a method to select key scales along phase singularity curves. We demonstrate two applications of PSs: object tracking and image matching. In object tracking, we use the iterative closest point algorithm to determine the correspondences of PSs between two adjacent frames. The use of PSs allows us to precisely determine the motions of tracked objects. In image matching, we combine PSs and scale-invariant feature transform (SIFT) descriptor to deal with the variations between two images and examine the proposed method on a benchmark database. The results indicate that our method can find more correct matching pairs with higher repeatability rates than some well-known methods.

Index Terms—Image matching, image representation, object tracking, phase singularity, scale space, transformation invariance.

I. INTRODUCTION

ONE of the fundamental problems in image processing and computer vision is image representation. A good representation should be compact and stable to noise addition, transformations, and image deformations, while providing rich and distinctive information for image processing and understanding

Manuscript received June 19, 2008; revised April 13, 2009. First published July 06, 2009; current version published September 10, 2009. Y. Qiao was supported in part by the Japan Society for the Promotion of Science (JSPS) (P07078). W. Wang and M. Takeda were supported by JSPS B(2) No.18360034. J. Liu and X. Tang were partly supported by a grant from the Research Grants Council of the Hong Kong SAR, China (Project No. CUHK 415408). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Minh N. Do.

Y. Qiao and N. Minematsu are with the Graduate School of Information Science and Technology, University of Tokyo, Tokyo, Japan (e-mail: qiao@gavo.t.u-tokyo.ac.jp; mine@gavo.t.u-tokyo.ac.jp).

W. Wang is with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, U.K. (e-mail: w.wang@hw.ac.uk).

J. Liu and X. Tang are with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China (e-mail: jzliu@ie.cuhk.edu.hk; xtang@ie.cuhk.edu.hk).

M. Takeda is with the Department of Information and Communication Engineering, University of Electro-Communications, Tokyo, Japan (e-mail: takeda@ice.uec.ac.jp).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2009.2026623

tasks. The phases, estimated by convolving images with complex filters, prove to be robust features with respect to noise additions and transformations [1]–[3]. Phases can smooth shading and brightness variation due to its invariant nature to amplitude [4]. Moreover, phases contain rich information of an input image. It is shown that one can recover the original image to a large extent by using phase information only (magnitudes are set to unity) [5]; on the other hand, if magnitudes are retained and phases are set to zero, the recovered image is completely indiscernible. Phases have found successes in many applications, such as disparity estimation of stereo [2], matching [4], [6], and face recognition [7].

Phase singularity (PS) refers to a point where a complex signal equals zero. In most previous phase-based work, PSs are generally regarded as unfavorable and unreliable positions [8]. The usual treatments fail at these positions as phases change abruptly around them and cannot be estimated. In [8] and [3], the authors analyzed the stability of phase to geometrical deformations, and concluded that the occurrence of PSs is one of the main causes of instability and PSs should be discarded.

Different from the classical view, in this paper, we find that PSs (estimated by using the Laguerre-Gauss filter) contain important information of an image and can provide an efficient tool for image representation and processing. Unlike direct use of phase information as in [4], [6], [8], and [3], our focus is on where and how PSs happen. We prove that PSs are invariant to translation and rotation, and find that the positions of PSs may contain nearly complete information for reconstructing the original image up to a scale. We study the core structure around a PS and introduce several invariant measures to characterize its core structure. We analyze the stability of PSs with noise addition and scale change. By using multiscale space analysis, we generalize PS points into PS curves. This allows us to select key PS points along the PS curves and determine their characteristic scales. We develop two applications of PSs, object tracking and image matching. In the first application, we determine the corresponding PSs on the tracked objects by using the iterative closest point (ICP) algorithm [9]. The use of PSs enables us to precisely determine the positions and the poses of the tracked objects. In image matching, we adopt the popular scalable invariant feature transform (SIFT) features [10] for robust PS matching. The results show that our method can find more correct matching pairs with higher repeatability rates than the well-known detector of Lowe [10]. A small portion of this study has been published in a conference paper [11]. Although the discussions and analysis are limited to the PSs estimated by the LG filter, many results

and methods in this paper can be generalized to other complex filters.

We also note that PSs have been intensively studied in physics and optics. PSs, also known as dislocations [12], defects [13], or optical vortices [14], prove to be an effective tool for describing and analyzing various physical phenomena such as rainbows, tide waves [15], nonlinear optics [14], condensed-matter [13], and black-body radiation [16]. More recently, Wang *et al.* used PSs to measure the displacement of speckle patterns, which is referred to as the Optical Vortex Metrology [17], [18].

II. PHASE SINGULARITY

Mathematically, phase is defined as the argument $\theta = \arctan(a/b)$ of a complex signal $a + ib$. Phase singularity refers to a point where the complex signal equals zero and the phase cannot be estimated. For a 2-D image $I(x, y)$, we can obtain its complex representation $\hat{I}(x, y)$ by convolving $I(x, y)$ with a 2-D complex filter $T(x, y) = T_r(x, y) + iT_i(x, y)$

$$\begin{aligned} \hat{I}(x, y) &= I(x, y) * T(x, y) \\ &= I(x, y) * T_r(x, y) + iI(x, y) * T_i(x, y) \end{aligned} \quad (1)$$

where “*” represents convolution operation.

In this paper, we adopt the Laguerre-Gauss (LG) filter [19]

$$LG(x, y, \sigma) = -\frac{x + yi}{2\pi\sigma^4} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (2)$$

where σ is a parameter of scale. This filter has been used in Wang *et al.*'s work on optical vortex metrology [17], [20]. Although we use the LG filter in this paper, many results can be applied or generalized to other complex filters, such as steerable filter [21], Gabor filter [22], [23], Hilbert transform [4], [5], and Hermite transform [24].

The real and imaginary parts of the LG filter correspond to the partial derivatives of a 2-D Gaussian function $G(x, y) = 1/(2\pi\sigma^2) \exp(-(x^2 + y^2)/(2\sigma^2))$, which have been widely used for edge detection and scale space analysis [25], [26]

$$\text{Re}\{LG(x, y)\} = G_x(x, y) \quad (3)$$

$$\text{Im}\{LG(x, y)\} = G_y(x, y). \quad (4)$$

Physiological evidence indicates that the profiles of visual receptive fields in the primate eye can be approximated by the Gaussian derivatives [27].

Let $E(x, y) = I(x, y) * G(x, y)$ denote the Gaussian smoothing image. We have $E_x(x, y) = I(x, y) * G_x(x, y)$ and $E_y(x, y) = I(x, y) * G_y(x, y)$. Thus, the complex image can be decomposed as

$$\hat{I}(x, y) = E_x(x, y) + iE_y(x, y). \quad (5)$$

The above formulation (5) indicates an efficient method to obtain \hat{I} by calculating the partial derivatives of $E(x, y)$. This can also be generalized to the higher order derivatives, for example

$$\begin{aligned} \hat{I}_x(x, y) &= E_{xx}(x, y) + iE_{yx}(x, y) \\ &= I(x, y) * G_{xx}(x, y) + iI(x, y) * G_{yx}(x, y). \end{aligned} \quad (6)$$

The Fourier transform of the LG filter (2) is given by

$$\mathbb{L}\mathbb{G}(f_x, f_y) = 2\pi(i f_x - f_y) \exp(-2\pi^2\sigma^2(f_x^2 + f_y^2)). \quad (7)$$

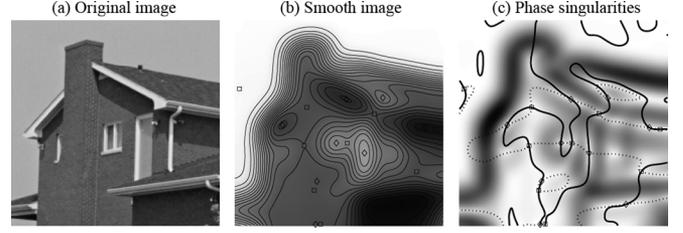


Fig. 1. (a) Original image. (b) Gaussian smoothing image $E(x, y)$ with level sets. The scale parameter $\sigma = 12$. (c) Filtered image $|\hat{I}(x, y)|$ with zero-crossing lines. The black lines in (b) denote the level-sets of $E(x, y)$. The diamonds and squares in (b) and (c) represent extreme and saddle PS points, respectively. The solid and dash lines in (c) represent the zero-crossing lines of the real and imaginary parts of $\hat{I}(x, y)$, respectively.

It can be seen from (7) that 1) the LG filter is a band-pass filter, and its bandwidth is determined by the scale parameter σ . This allows us to control the density (number) of PSs in $\hat{I}(x, y)$ by changing σ ; 2) the LG filter is DC free, which makes it robust to homogeneous illumination variations; 3) the LG filter is “self-Fourier” in the sense that it has the same functional form as its 2-D Fourier transformation (only certain parameters are changed or inverted).

As another choice for calculating (1), we can calculate the Fourier transform of $I(x, y)$ first, then multiply it with $\mathbb{L}\mathbb{G}(f_x, f_y)$, and finally obtain the complex image $\hat{I}(x, y)$ by taking the inverse Fourier transform of the multiplication.

PS points, or PSs for short, are the positions where $\hat{I}(x, y) = 0$. Formally, we use $PS(I, \sigma) = \{(x, y) | \hat{I}(x, y, \sigma) = 0\}$ to denote the set of PSs. PSs can also be seen as the intersection points of two zero-crossing curves, $E_x(x, y) = 0$ and $E_y(x, y) = 0$. An example of a LG-filtered image and the associated PSs are shown in Fig. 1. An important property of PSs are their invariance to certain transformations described by the following theorem.

Theorem 1: The translated and rotated version of an image has the same PSs as those in the original image.

The proof of Theorem 1 is straightforward. Although the invariance does not hold for image scaling, we will analyze how PS points are affected by image scaling in Section III-A. The image scaling problem can be further solved in a solid framework through key PS selection in Section III-C. Small noise addition or image distortion can change the positions of PS points. However, they are topologically-stable points, and small perturbation usually cannot eliminate or create PS points [28]. We analyze the stability of PS points to noise addition in Section II-C.

A. Core Structure of PSs

In this section, we study the core structure of PSs and introduce several invariant measures to characterize their core structures. These measures have their roots in physical studies [16], [17], and are invariant to certain transformations. We also discuss the relation among different types of PSs, and show how PSs are related to level-sets of Gaussian smoothing image and zero-crossing lines of complex image.

In this paper, we mainly consider PSs with nonsingular Hessian matrices. These PSs can be divided into two classes: *extremes* and *saddles*, which correspond to the extremal and saddle points of $E(x, y)$, respectively (shown as diamonds and squares in Fig. 1). Examples of extreme and saddle points are given in

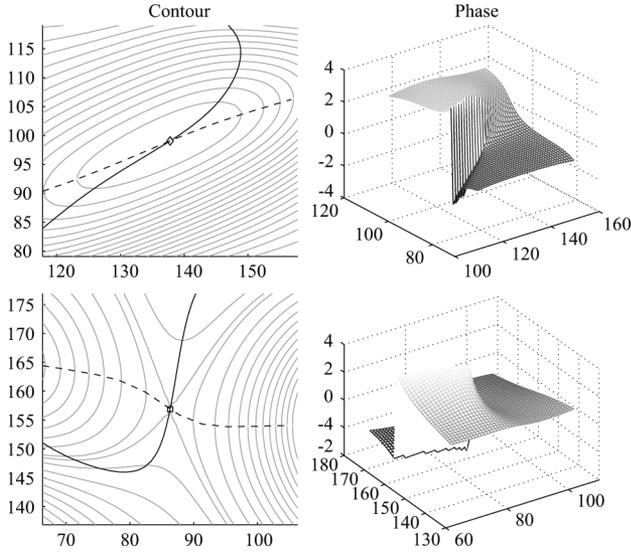


Fig. 2. Core structure of PS points. (a) An extreme point (137.1, 99.8) in Fig. 1(b). (b) A saddle point (86.3, 156.9) in Fig. 1(b). Gray lines represent the level-sets of $E(x, y)$. The solid lines and dash lines represent the real and imaginary zero-crossing lines of $\hat{I}(x, y)$, respectively.

Fig. 1(b) and (c). For convenience, we call the lines with constant density in the Gaussian smoothing image $E(x, y)$ as *level sets*, and the lines with zero real (imaginary) parts of $\hat{I}(x, y)$ as *real (imaginary) zero-crossing lines*. Each extreme point is encircled by a set of closed level-sets [Figs. 1(b) and 2(a)]. A saddle point is the self-intersection of a level-set [Figs. 1(b) and 2(b)]. PSs also correspond to intersection points of real and imaginary zero-crossing lines [Figs. 1(c) and 2]. An extreme point corresponds to a local maximum or minimum value of $E(x, y)$ with the two zero-crossing lines intersecting at it, while a saddle point has a local maximum value in one zero-crossing line and a minimum value in another.

It is easy to see that the Jacobian matrix of the real and imaginary parts of $\hat{I}(x, y)$ equals the Hessian matrix of $E(x, y)$

$$\begin{aligned} J(\hat{I}(x, y)) &= \begin{bmatrix} \text{Re}\{LG(x, y)\}_x & \text{Re}\{LG(x, y)\}_y \\ \text{Im}\{LG(x, y)\}_x & \text{Im}\{LG(x, y)\}_y \end{bmatrix} \\ &= \begin{bmatrix} E_{xx} & E_{xy} \\ E_{yx} & E_{yy} \end{bmatrix} = H(E(x, y)). \end{aligned} \quad (8)$$

The above matrix is also called *vorticity* in physics [16]

$$\begin{aligned} \Omega(p) &= \nabla \text{Re}\{LG(x, y)\} \times \nabla \text{Im}\{LG(x, y)\} \\ &= J(\hat{I}(x, y)) = H(E(x, y)). \end{aligned} \quad (9)$$

The sign of vortex p is defined as the sign of the determinant of vorticity matrix [16], [29]: $s(p) = \text{sgn}(E_{xx}E_{yy} - E_{xy}^2)$. In this sense, extreme points are positive and saddle points are negative. The *topological charge* of p is defined as

$$c(p) = \frac{1}{2\pi} \oint_C d\theta(x, y) \quad (10)$$

where C is a counter-clockwise closed curve enclosing p . The topological charge is positive when the phase increases along C . $c(p)$ must be an integer. Recall that

$\theta(x, y) = \arctan(E_y(x, y)/E_x(x, y))$. For an extreme point, $\theta(x, y)$ increases along C except at the abrupt change from π to $-\pi$ [Fig. 2(a)]; while for a saddle point, $\theta(x, y)$ decreases except at the abrupt change from $-\pi$ to π [Fig. 2(b)]. Thus, $c(p)$ has the same sign with $s(p)$. In fact, the topological charge is always +1 at an extreme point. At a saddle point, if the number of zero-crossing lines connected with it is $2k$, its topological charge is $1 - k$, where k is a positive integer. The signs of neighboring PSs satisfy the following theorem [29].

Theorem 2: Two adjacent PSs along a zero-crossing (real or imaginary) line have opposite signs.

The above theorem can be verified in Fig. 1(c), where a diamond (an extreme point) always has squares (saddle points) as its neighbors, and vice versa. One can refer to [30] for an extensive study on the topological relations among saddles, extremes and level-sets. [16] describes a solid theoretical study on the geometry and statistics of PS points, such as density, curvature and correlation. Their results indicated that the statistics of PSs depend on the frequency spectrum of $E(x, y)$.

We use the directions of zero-crossing lines θ_1 and θ_2 to describe the orientation of a PS point. For a real zero-crossing line (such as the solid line in Fig. 2)

$$\theta_1 = \arctan\left(\frac{dx}{dy}\right) = -\arctan\left(\frac{E_{xy}}{E_{xx}}\right) \quad (11)$$

and for an imaginary zero-crossing line (such as the dashed line in Fig. 2)

$$\theta_2 = \arctan\left(\frac{dx}{dy}\right) = -\arctan\left(\frac{E_{yy}}{E_{xy}}\right). \quad (12)$$

The angle between two zero-crossing lines is calculated by [17], [18]

$$\theta_{RI} = |\theta_1 - \theta_2|. \quad (13)$$

The real and imaginary parts of a complex signal near a PS can be approximated by two planes [17], [20]

$$\text{Re}\{\hat{I}(x, y)\} \approx a_r x + b_r y + c_r \quad (14)$$

$$\text{Im}\{\hat{I}(x, y)\} \approx a_i x + b_i y + c_i. \quad (15)$$

In this way, the level-sets of $|\hat{I}(x, y)|$ around a PS can be approximated by a set of concentric ellipses, whose eccentricity is estimated by [17], [20]: see (16), shown at the bottom of the next page.

It is not difficult to verify that vortex's sign s , vorticity determinant $|\Omega|$, crossing angle θ_{RI} , and eccentricity e are invariant to the in-plane transformations of rotations and translations [17], [18].

B. Image Reconstruction

PS points contain nearly complete information of an image. In [31], Kanters *et al.* developed a method to reconstruct an image from multiscale critical points (corresponding to PS points). However, they required the derivatives at critical points for reconstruction, and did not discuss that under what conditions the image can be reconstructed. Here we show that generally it is possible to reconstruct an image up to a scale if enough "independent" PS points $\{p_m(x_m, y_m, \sigma_m)\}$ can be obtained. Sup-

pose that $I(x, y)$ can be expanded into a polynomial up to order n and

$$I(x, y) = \sum_{\substack{i,j \\ (n \geq i+j \geq 0)}} a_{i,j} x^i y^j. \quad (17)$$

The problem of recovering $I(x, y)$ is then equivalent to determining the coefficients $a_{i,j}$'s. By convolving the image with the LG filter, we have

$$\begin{aligned} E_x &= I(x, y) * G_x(x, y) \\ &= \iint \sum_{\substack{i,j \\ (n \geq i+j \geq 0)}} a_{i,j} (x-u)^i (y-v)^j \\ &\quad \times G_x(u, v) dudv \end{aligned} \quad (18)$$

and

$$\begin{aligned} E_y &= I(x, y) * G_y(x, y) \\ &= \iint \sum_{\substack{i,j \\ (n \geq i+j \geq 0)}} a_{i,j} (x-u)^i (y-v)^j \\ &\quad \times G_y(u, v) dudv. \end{aligned} \quad (19)$$

Since

$$\begin{aligned} G^n(\sigma) &= \int_{-\infty}^{\infty} \frac{x^n}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\ &= \begin{cases} (n-1)!!\sigma^n, & n \text{ is even} \\ 0, & n \text{ is odd} \end{cases} \end{aligned} \quad (20)$$

where $n!!$ is the double factorial defined by

$$n!! = \begin{cases} n(n-2)\dots 3 \cdot 1, & n \text{ is odd} \\ n(n-2)\dots 4 \cdot 2, & n \text{ is even} \\ 1, & n = -1, 0 \end{cases} \quad (21)$$

we have

$$\begin{aligned} M_{i,j}^1(x, y, \sigma) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-u)^i (y-v)^j G_x(u, v) dudv \\ &= \int_{-\infty}^{\infty} \frac{-u}{\sqrt{2\pi}\sigma^3} \exp\left(-\frac{u^2}{2\sigma^2}\right) (x-u)^i du \\ &\quad \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{v^2}{2\sigma^2}\right) (y-v)^j dv \\ &= -\frac{1}{\sigma^2} \left(\sum_{k=0}^i {}_i C_k (-1)^{i-k} x^k G^{i-k+1}(\sigma) \right) \\ &\quad \cdot \left(\sum_{k=0}^j {}_j C_k (-1)^{j-k} y^k G^{j-k}(\sigma) \right) \end{aligned} \quad (22)$$

$$\begin{aligned} M_{i,j}^2(x, y, \sigma) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-u)^i (y-v)^j G_y(u, v) dudv \\ &= \int_{-\infty}^{\infty} \frac{-1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{u^2}{2\sigma^2}\right) (x-u)^i du \\ &\quad \cdot \int_{-\infty}^{\infty} \frac{v}{\sqrt{2\pi}\sigma^3} \exp\left(-\frac{v^2}{2\sigma^2}\right) (y-v)^j dv \\ &= -\frac{1}{\sigma^2} \left(\sum_{k=0}^i {}_i C_k (-1)^{i-k} x^k G^{i-k}(\sigma) \right) \\ &\quad \cdot \left(\sum_{k=0}^j {}_j C_k (-1)^{j-k} y^k G^{j-k+1}(\sigma) \right). \end{aligned} \quad (23)$$

It is impossible to uniquely reconstruct image $I(x, y)$ from PSs. This is because $I(x, y)$ and $\alpha I(x, y)$ ($\alpha \neq 1$ is a constant) have the same PSs. However, we can determine $b_{i,j} = a_{i,j}/a_{0,0}$ from enough PSs, where the number of variables $\{b_{i,j}\}$ is $(n+1)(n+2)/2 - 1$. In an extreme case, for a completely flat image with constant density, every pixel becomes a PS. We cannot estimate the pixel density from PSs. For every PS point $p_m(x_m, y_m, \sigma_m)$, we can obtain two linear equations for $b_{i,j}$

$$\begin{aligned} \sum_{\substack{i,j \\ (n \geq i+j \geq 1)}} M_{i,j}^1(x_m, y_m, \sigma_m) b_{i,j} \\ = -M_{0,0}^1(x_m, y_m, \sigma_m) \end{aligned} \quad (24)$$

$$\begin{aligned} \sum_{\substack{i,j \\ (n \geq i+j \geq 1)}} M_{i,j}^2(x_m, y_m, \sigma_m) b_{i,j} \\ = -M_{0,0}^2(x_m, y_m, \sigma_m). \end{aligned} \quad (25)$$

Use two column vectors v_m^1 and v_m^2 to represent the coefficients $M_{i,j}^1(x_m, y_m, \sigma_m)$ and $M_{i,j}^2(x_m, y_m, \sigma_m)$, $n \geq i+j \geq 1$, respectively. Then we only need to find a certain number of PS points $\{p_m\}_{m=1}^M$, such that the matrix $V = [v_1^1, v_1^2, \dots, v_M^1, v_M^2]^T$ has a rank of $(n+1)(n+2)/2 - 1$. This results in a sufficient number of equations to solve for $\{b_{i,j}\}$. It is easy to see that the number of points is at least $\lceil (n^2 + 3n)/4 \rceil$.

The above analysis indicates that PSs, the abnormal points, can provide a nearly complete description of an image up to a scale. In spite of the polynomial form, we can also reconstruct the image based on the coefficients of discrete Fourier transformation (DFT). The details are omitted due to space limitation.

C. Stability of PSs to Noise

Small perturbations usually do not create new PS points or eliminate existing ones, since PS points are individually topo-

$$e = \sqrt{1 - \frac{a_r^2 + a_i^2 + b_r^2 + b_i^2 - ((a_r^2 + a_i^2 - b_r^2 - b_i^2)^2 + 4(a_r b_r + a_i b_i)^2)^{1/2}}{a_r^2 + a_i^2 + b_r^2 + b_i^2 + ((a_r^2 + a_i^2 - b_r^2 - b_i^2)^2 + 4(a_r b_r + a_i b_i)^2)^{1/2}}} \quad (16)$$

logically stable [28], [32], but they can cause the changes of the positions of PSs. In this section, we study the stability of PSs to noise addition. It should be mentioned that our stability analysis is different from that in [3]. Their focus is on the changes of phases, while ours is on the changes of PSs' positions.

Let $N(x, y)$ be the noise added to image $I(x, y)$, and $p(x, y)$ be a PS point in $I(x, y)$. Then the noisy image is $I'(x, y) = I(x, y) + N(x, y)$, and PS $p(x, y)$ moves to a new position denoted by $p'(x', y')$ in $I'(x, y)$. Our objective is to estimate the gradient vector $\nabla_{\mathbf{x}} = [dx, dy]^T$, where $dx = x' - x$ and $dy = y' - y$ are the displacements. Applying the LG filter to $I'(x, y)$, we have

$$\begin{aligned} \hat{I}'(x, y) &= (I(x, y) + N(x, y)) * LG(x, y) \\ &= \hat{I}(x, y) + \hat{N}(x, y). \end{aligned} \quad (26)$$

Assume that $N(x, y)$ is i.i.d. (independent and identically distributed) additive Gaussian noise with mean μ and standard deviation Σ . We take $\hat{N} = \hat{N}_r + i\hat{N}_i$ as a 2-D vector $[\hat{N}_r, \hat{N}_i]^T$. Then \hat{N} can be described by a 2-D Gaussian distribution with zero mean and the following diagonal covariance matrix

$$\Sigma_{\hat{N}} = \Sigma^2 \begin{bmatrix} \frac{1}{8\pi\sigma^4} & 0 \\ 0 & \frac{1}{8\pi\sigma^4} \end{bmatrix}. \quad (27)$$

Using Taylor expansion on $\text{Re}\{\hat{I}(x', y')\}$ and $\text{Im}\{\hat{I}(x', y')\}$, and omitting the terms with orders higher than 2, we have

$$\begin{aligned} \text{Re}\{\hat{I}(x', y')\} &\approx \nabla_{\mathbf{x}}^T \begin{pmatrix} E_{xx} \\ E_{xy} \end{pmatrix} \\ &+ \nabla_{\mathbf{x}}^T \begin{pmatrix} E_{xxx} & E_{xxy} \\ E_{xxy} & E_{xyy} \end{pmatrix} \nabla_{\mathbf{x}} \end{aligned} \quad (28)$$

$$\begin{aligned} \text{Im}\{\hat{I}(x', y')\} &\approx \nabla_{\mathbf{x}}^T \begin{pmatrix} E_{xy} \\ E_{yy} \end{pmatrix} \\ &+ \nabla_{\mathbf{x}}^T \begin{pmatrix} E_{xxy} & E_{xyy} \\ E_{xyy} & E_{yyy} \end{pmatrix} \nabla_{\mathbf{x}}. \end{aligned} \quad (29)$$

Then we can approximate the probability of point (x', y') being a PS in image \hat{I} by (30), shown at the bottom of the page. We can insert (28) and (29) into (30) to get an explicit form. Note that with only the first order of (28) and (29), the probability of (30) reduces to a Gaussian distribution of $\nabla_{\mathbf{x}}$.

We also conduct experiments to examine the shift of PSs caused by noise addition with the image set constructed in [33], which includes images of various contents and textures. This set has been widely used for evaluating image matching, and is adopted in our experiment 2 (Section V). For each image, we add Gaussian noise to it and calculate two sets of PSs in the original image I and the noisy image I' . Then the two PS sets are compared by two criteria: 1) shift error: the distance between

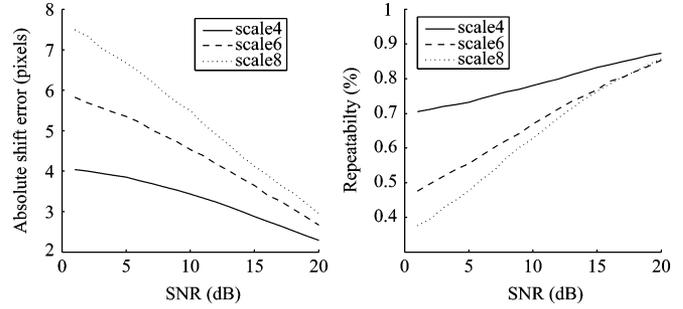


Fig. 3. Average absolute shift error and repeatability of PSs versus noise addition with different scales. The total numbers of PSs are 214077, 99701, and 57078, for scales 4, 6, and 8, respectively.

each PS in I and its nearest PS in I' ; 2) repeatability: the percentage of PSs in I , which are repeated in I' with shift of at most 5 pixels. We repeat this procedure for each image and each noise level. The average results over all images are shown in Fig. 3. As expected, the shift error linearly decreases as SNR increases. It can also be found that the shift error increases as scale σ increases. This is because \hat{I} becomes flatter as scale increases, which makes it more sensitive to additive noise. However, it should be noted that this does not mean that the smaller the scale is, the more desirable it is. In fact, a small scale corresponds to a small image patch, which is sensitive to image deformation. The number of PSs increases exponentially with the decrease of the scale, and it can be equal to or more than the number of pixels in the image if the scale is small enough. A large number of PSs not only lead to extensive computation but also diminish the compactness for using PSs to represent an image. We will discuss the scale selection problem in Section III-C. In the experiments (Sections IV and V), we can see that the PSs are generally stable to real noise and image deformation.

III. SCALE SPACE ANALYSIS OF PHASE SINGULARITY CURVES

In Section II, we study the properties of PSs in a fixed scale and show its invariance to translation and rotation. The calculation of PSs depends on the scale parameter σ of LG transformation. Different settings of the scale parameter lead to different results: new PS points may appear, and existing ones may move or disappear. The bandwidth of the LG filter increases as σ decreases. Usually, the larger σ is, the fewer PS points we have. So we have to solve the problem of how to select good scales for the calculation of PSs. However, a single scale is seldom sufficient. In fact, the objects in an image may locate at different distances from the camera and have different sizes (corresponding to different scales). It is difficult to assert that one scale is better than another. Therefore, other than using a single scale, we consider PSs in a scale space coordinated by (x, y, σ) . The scale

$$\begin{aligned} p(\hat{I}'_x(x', y') = 0, \hat{I}'_y(x', y') = 0) \\ = \frac{1}{2\pi|\Sigma_{\hat{N}}|^{1/2}} \exp \left\{ -\frac{(\text{Re}\{\hat{I}(x', y')\})^2 + (\text{Im}\{\hat{I}(x', y')\})^2}{2|\Sigma_{\hat{N}}|} \right\} \end{aligned} \quad (30)$$

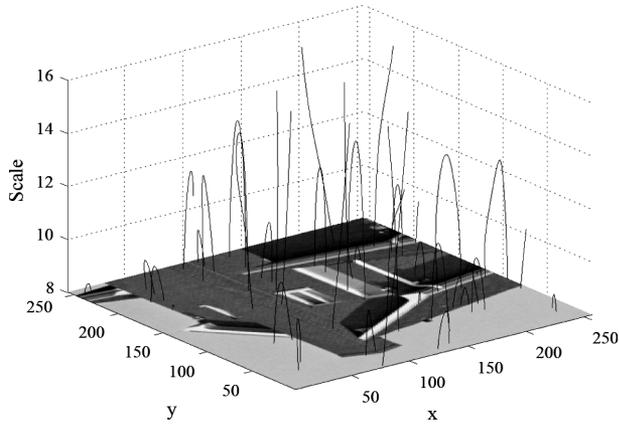


Fig. 4. Examples of PS curves in scale space.

space analysis was first studied by Iijima [34] in 1960s, and then became well known due to the works of Witkin [35] and Koenderink [36]. Lindeberg [37] provided a good introduction of scale space analysis for computer vision. In our scale space analysis, PS points are connected into curves, named *PS curves* as shown in Fig. 4. The positions of PSs are not invariant to scale change. As we show in the next, the scale space analysis can help to solve this problem.

A. Stability of PSs to Scale Changes

This section studies how PSs move when the scale changes. In 3-D scale space, PS points connect each other into PS curves. Let $p(x, y, \sigma)$ denote a point in a PS curve $L = (x(t), y(t), \sigma(t))$, where t is a curve parameter. L can be seen as an intersection curve of two 3-D surfaces, denoted by

$$E_x(x, y, \sigma) = 0 \quad (31)$$

$$E_y(x, y, \sigma) = 0. \quad (32)$$

From these two equations, we have

$$\frac{dE_x}{dt} = \frac{\partial E_x}{\partial x} \frac{dx}{dt} + \frac{\partial E_x}{\partial y} \frac{dy}{dt} + \frac{\partial E_x}{\partial \sigma} \frac{d\sigma}{dt} = 0 \quad (33)$$

$$\frac{dE_y}{dt} = \frac{\partial E_y}{\partial x} \frac{dx}{dt} + \frac{\partial E_y}{\partial y} \frac{dy}{dt} + \frac{\partial E_y}{\partial \sigma} \frac{d\sigma}{dt} = 0. \quad (34)$$

To estimate the movement of PSs with scale change, we calculate the derivatives $dx/d\sigma$ and $dy/d\sigma$ along L . Setting $t = \sigma$ and solving (33) and (34) ($E_{xy} = E_{yx}$), we have

$$\frac{dx}{d\sigma} = - \frac{E_{x\sigma}E_{yy} - E_{y\sigma}E_{xy}}{E_{xx}E_{yy} - E_{xy}^2} \quad (35)$$

$$\frac{dy}{d\sigma} = - \frac{E_{y\sigma}E_{xx} - E_{x\sigma}E_{xy}}{E_{xx}E_{yy} - E_{xy}^2}. \quad (36)$$

Recall that the Gaussian smoothing image $E(x, y)$ satisfies the following *diffusion equation* [37], [38]

$$E_{xx} + E_{yy} = \frac{2}{\sigma} E_{\sigma} \quad (37)$$

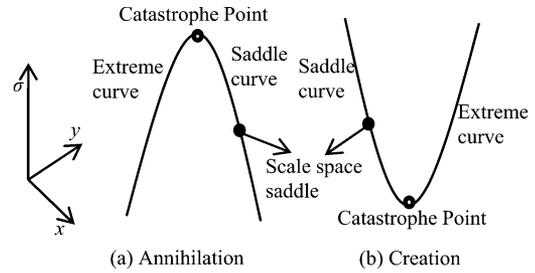


Fig. 5. Annihilation, creation, and PS curves. Hessian determinant $|\Omega| = 0$ at a catastrophe point, $|\Omega| > 0$ along an extreme curve, $|\Omega| < 0$ along a saddle curve, and $\Delta E = 0$ at a space scale saddle.

where E_{σ} denotes the partial derivative of $E(x, y)$ with respect to scale σ . Thus, we have

$$E_{x\sigma} = \frac{\sigma}{2} (E_{xxx} + E_{yyx}) \quad (38)$$

$$E_{y\sigma} = \frac{\sigma}{2} (E_{xxy} + E_{yyy}). \quad (39)$$

The above equations allow us to estimate the stability of PSs to scale changes. It can be found that the denominators of (35) and (36) consist of the determinant of vorticity $|\Omega|$ (see (9)). Roughly speaking, the larger $|\Omega|$ is, the smaller $dx/d\sigma$ and $dy/d\sigma$ are, and the more stable the PSs are. Also one can find in Fig. 4 that most PS-curves are nearly vertical, which indicates that mostly the derivatives of (35) and (36) are small and the associated PSs are stable to scale change.

B. Properties of PS Curves

In catastrophe theory [39] and topological structure analysis of scale space [40], PSs with nonzero $|\Omega|$ correspond to Morse critical points, PSs with zero $|\Omega|$ are called non-Morse points, catastrophe points or toppoints, and PS curves correspond to critical curves. According to the Morse lemma [38] and Thom theorem [41], the neighborhoods of a Morse critical point can be described by a second order polynomial in canonical coordinates, while the local topology of a toppoint requires a higher order polynomial. A classification study of toppoints was described in [42]. There are two basic generic events in Gaussian scale space: annihilations and creations of one extreme and one saddle. Damon [38] proved that all the other events can be reduced into a combination of one of these two events and one “in which nothing happens”.

Like PS points, PS curves can also be divided into two types: extreme curves and saddle curves. Extreme curves can be further classified into maximum curves and minimum curves. An extreme curve is jointed with a saddle curve at a catastrophe point where annihilation or creation happens as shown in Fig. 5. The PSs along an extreme curve and a saddle curve have different signs of $|\Omega|$, and at a catastrophe point, $|\Omega| = 0$. This indicates that one can trace a PS curve until meeting a catastrophe point. According to catastrophe theory, each PS curve must have its upper bound of the scale. If the scale is large enough, there is only one extreme left. An extreme curve satisfies the following *extreme principle* [37].

Theorem 3: The Gaussian smoothing density $E(x, y)$ decreases (increases) along a maximum (minimum) curve when the scale increases.

The above theorem states the nonenhancement of local extremes in Gaussian scale space. The proof is easy: at a local maximum, Laplacian $\Delta E = E_{xx} + E_{yy}$ is negative, while at a local minimum, ΔE is positive. From (37), the Laplacian has the same sign as the partial derivative E_σ . Along a saddle curve, generally there exists a point where the Laplacian $\Delta E = 0$, called scale space saddle. Since the sign of ΔE on the two sides of a scale space saddle is different, the density E changes oppositely. This means that a scale space saddle corresponds to a local maximum (minimum) of the density E along a saddle curve. Kuijper found the Iso-Intensity Manifolds through saddle points provide a scale space hierarchy tree representation and used this for image segmentation [43], [44].

C. Key PS Point Selection

It can be seen from Fig. 4 that PS points of different scales along a PS curve are highly correlated; their x and y coordinates are usually similar, since PS curves are nearly vertical. The direct use of PS curves for image representation is obviously very redundant. Therefore, it is necessary to study scale (PS point) selection problem along PS curves. An image usually contains objects of different sizes. The intrinsic problem is to determine the *characteristic scale* of PS points, from which we can estimate the “physical length” of objects (or structures). It is desirable if we can select scales according to the sizes of the objects. However, in most cases, we do not have prior knowledge of the sizes of the objects. To circumvent this difficulty, we adapt the principle proposed by Lindeberg [25], which states that the scale level, where a certain combination of normalized derivatives [refer to (42) below] has a local maximum, reflects a characteristic scale length of a corresponding structure in an image. For example, 1) for a sinusoidal signal, the normalized derivatives achieve maximum at the scales proportional to the wavelength of the signal; 2) the normalized Laplacians have maximum at scales corresponding to the sizes of blob patterns.

In the following, we show how the image scaling problem can be solved by key PS selection. Consider two images I and I' related by scaling

$$I(x, y) = I'(sx, sy) \quad (40)$$

where s is a zooming factor. Let $x' = sx, y' = sy$, and $\sigma' = s\sigma$. Then $I(x, y) = I'(x', y')$ and $E(x, y, \sigma) = E'(x', y', \sigma')$, where E' is the Gaussian smoothing image of I' . It is not difficult to examine that the derivatives of E and E' satisfy

$$E_{x^m y^n}(x, y, \sigma) = s^{m+n} E'_{x^m y^n}(x', y', \sigma') \quad (41)$$

where m and n denote the orders of derivative. Therefore, $\hat{I}(x, y, \sigma) = s\hat{I}'(x', y', \sigma')$.

The normalized derivatives are defined by [25]

$$E_{x^m y^n}^{\text{norm}}(x, y, \sigma) = \sigma^{m+n} E_{x^m y^n}(x, y, \sigma). \quad (42)$$

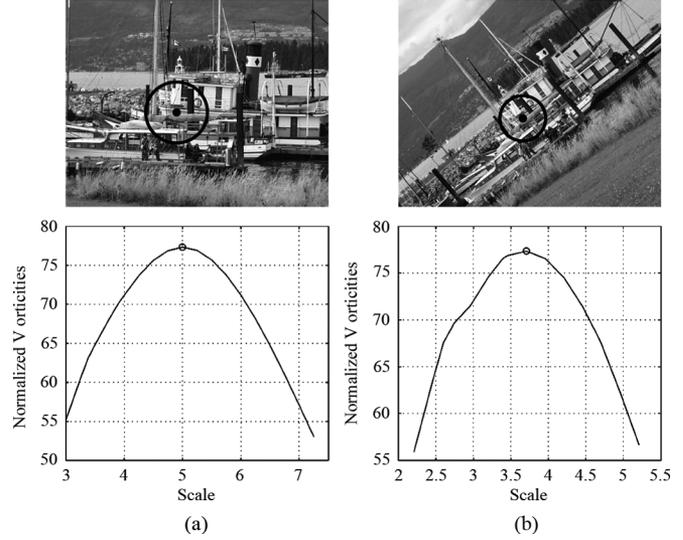


Fig. 6. Example of PS selection by maximizing the normalized vorticity (NV) in two matching images. The two black points in the upper figures denote two key PSs. And the radius of the circles equal to 20 times of the key scales of the two PSs. One can find that the two circles have almost the same contents. The two figures below show the NV functions of the two PS curves which pass through the two PSs.

The following are two examples:

$$\begin{aligned} \text{Normalized Laplacian : } NL(x, y, \sigma) \\ = \sigma^2(E_{xx} + E_{yy}) \end{aligned} \quad (43)$$

$$\begin{aligned} \text{Normalized Vorticity : } NV(x, y, \sigma) \\ = \sigma^4 \Omega(x, y, \sigma). \end{aligned} \quad (44)$$

Then it can be found that

$$E_{x^m y^n}^{\text{norm}}(x, y, \sigma) = E_{x^m y^n}^{\text{norm}}(x', y', \sigma'). \quad (45)$$

The above equation yields an invariant measure for images before and after zooming. Consider two corresponding PS curves $PS(t) = (x(t), y(t), \sigma(t))$ and $PS'(t') = (x'(t'), y'(t'), \sigma'(t'))$ of \hat{I} and \hat{I}' , where $t' = st$. The normalized derivatives $E_{x^m y^n}^{\text{norm}}$ and $E_{x^m y^n}^{\text{norm}}$ have the same values at two corresponding points $PS(t)$ and $PS'(t')$. Thus, PS and PS' achieve their local extremes (maximum or minimum) at the corresponding positions and scales. Let (x_M, y_M, σ_M) and (x'_M, y'_M, σ'_M) denote two corresponding extremes of E along the PS curves. Then $E_{x^m y^n}^{\text{norm}}(x_M, y_M, \sigma_M) = E_{x^m y^n}^{\text{norm}}(x'_M, y'_M, \sigma'_M)$. An example is shown in Fig. 6. We can estimate the zooming factor by

$$s = \frac{\sigma'_M}{\sigma_M}. \quad (46)$$

The above analysis lead to a method to select key PS points along PS curves by maximizing (minimizing) the normalized derivatives. This has important practical significance. For example, in image matching and registration, we want to determine the matching points between two images, which may undertake transformations. We have shown that PS points are invariant to rotation and translation. By searching for extremes

along PS curves, we can find key PS points with corresponding scales in the two images. This allows us not only to solve the image scaling problem but also to determine the characteristic scales associated with the key PSs.

The main difference between our method and previous scale selection methods [10], [33], [45] is that our selection is constrained to PS curves. Unlike previous selection methods, we do not need to compare the normalized Laplacian of every point with its neighbors in scale space for selecting extreme points. Moreover, the use of PS curves allows us to achieve sub-pixel precision.

IV. APPLICATION 1: OBJECT TRACKING

PSs are expected to have a number of potential applications in image processing and computer vision. We illustrate two applications in this paper. This section shows how PSs can be applied on object tracking. In the context of object tracking, the positions of the object in the previous frame I_{t-1} is available and the objective is to determine the object's position in the current frame I_t . The problem of object tracking has been under intensive studies in computer vision. The tracking methods differ in the models of objects and the assumptions/constraints on motions and/or cameras. A comprehensive survey on object tracking can be found in [46]. Although many tracking methods can deal with large image deformations and abrupt motions, one of the common shortcomings of these methods is that they only provide relatively rough positions of the tracked object. Since PS calculation can provide sub-pixel locations of characteristic points, we can develop a precise tracking method based on PSs.

We calculate a set of PS points on a tracked object in the previous frame I_{t-1} . Then the tracking problem becomes identifying their corresponding PS points in the current frame I_t . Because the sizes of the object in two adjacent frames do not change much, we do not consider image scaling problem in PS calculation. Perhaps the simplest idea to determine the matching of PSs is to find the corresponding pairs with the least difference of the local core measures, such as topological charge c , vorticity Ω , eccentricity e , and crossing angle θ_{RI} introduced in Section II-A. However, we found that this method leads to miss-matchings due to complex textures and noise in images. For this reason, we only use these local core features to find the candidate matching PSs, and then determine the final matching pairs by the spatial structure (geometrical information) of PSs. The spatial structure refers to the mutual position relations between the individual PSs, which remains stable between two adjacent frames. Thus, the spatial structure can be used for the unique identification of the corresponding relations of PSs.

The details are formally described as follows. Let $G_{t-1} = \{PS_{t-1}^i\}_{i=1}^n$ and $G_t = \{PS_t^j\}_{j=1}^m$ denote two groups of PS points on the tracked object in I_{t-1} and I_t , respectively. Our objective is to determine the corresponding relations $j = M(i)$ between G_{t-1} and G_t , where PS_{t-1}^i and PS_t^j are two matching points ($M(i) = 0$ if the corresponding point does not exist).

For each PS_{t-1}^i , firstly, we identify the set $D(i)$ of its candidate corresponding points using the local core measures introduced in Section II-A with certain thresholds [17], [18]

$$D(i) = \left\{ j \mid \begin{aligned} &c(PS_{t-1}^i) = c(PS_t^j) \\ &\left| \frac{(\Omega(PS_{t-1}^i) - \Omega(PS_t^j))}{(\Omega(PS_{t-1}^i) + \Omega(PS_t^j))} \right| < \Omega_{th} \\ &|\theta_{RI}(PS_{t-1}^i) - \theta_{RI}(PS_t^j)| < \theta_{th} \\ &|e(PS_{t-1}^i) - e(PS_t^j)| < e_{th} \end{aligned} \right\} \quad (47)$$

where Ω_{th} , θ_{th} , and e_{th} are three thresholds.

Let (x_{t-1}^i, y_{t-1}^i) and (x_t^j, y_t^j) denote the coordinates of PS_{t-1}^i and PS_t^j . Using structural representation of PSs, the geometrical matching of two groups of PSs is related to determining the geometrical transformation between them. Let $T = (s, \beta, t_x, t_y)$ denote the transformation parameters, where s represents scaling, β rotation, and (t_x, t_y) translation. The ideal transformation of two matching points PS_{t-1}^i and PS_t^j is given by

$$\begin{pmatrix} x_t^j \\ y_t^j \end{pmatrix} = \begin{pmatrix} s \cos \beta & -s \sin \beta \\ s \sin \beta & s \cos \beta \end{pmatrix} \begin{pmatrix} x_{t-1}^i \\ y_{t-1}^i \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}. \quad (48)$$

Then the optimal matching with respect to minimum squared error can be formulated as

$$\min_{M, T} \sum_{M(i) \neq 0} \left\| \begin{pmatrix} s \cos \beta & -s \sin \beta \\ s \sin \beta & s \cos \beta \end{pmatrix} \begin{pmatrix} x_{t-1}^i \\ y_{t-1}^i \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} - \begin{pmatrix} x_t^{M(i)} \\ y_t^{M(i)} \end{pmatrix} \right\|^2. \quad (49)$$

In (49), if the transformation T is known, we can find the optimal matching by bipartite matching or nearest neighbor methods. On the other hand, if the matching M is known, we can determine the optimal transformation parameters as

$$s = \frac{\sqrt{A^2 + B^2}}{C} \quad (50)$$

$$\beta = \arctan \frac{B}{A} \quad (51)$$

$$t_x = \bar{x}_t - s(\cos \beta) \bar{x}_{t-1} + s(\sin \beta) \bar{y}_{t-1} \quad (52)$$

$$t_y = \bar{y}_t - s(\sin \beta) \bar{x}_{t-1} - s(\cos \beta) \bar{y}_{t-1} \quad (53)$$

where

$$\bar{x}_t = \frac{1}{n} \sum_{M(i) \neq 0} x_t^i \quad (54)$$

$$\bar{y}_t = \frac{1}{n} \sum_{M(i) \neq 0} y_t^i \quad (55)$$

$$\bar{x}_{t-1} = \frac{1}{n} \sum_{M(i) \neq 0} x_{t-1}^{M(i)} \quad (56)$$

$$\bar{y}_{t-1} = \frac{1}{n} \sum_{M(i) \neq 0} y_{t-1}^{M(i)} \quad (57)$$

$$A = \sum_{M(i) \neq 0} \left\{ (x_{t-1}^i - \bar{x}_{t-1})(x_t^{M(i)} - \bar{x}_t) + (y_{t-1}^i - \bar{y}_{t-1})(y_t^{M(i)} - \bar{y}_t) \right\} \quad (58)$$

$$B = \sum_{M(i) \neq 0} \left\{ (x_{t-1}^i - \bar{x}_{t-1})(y_t^{M(i)} - \bar{y}_t) - (y_{t-1}^i - \bar{y}_{t-1})(x_t^{M(i)} - \bar{x}_t) \right\} \quad (59)$$

$$C = \sum_{M(i) \neq 0} \left\{ (x_{t-1}^i - \bar{x}_{t-1})^2 + (y_{t-1}^i - \bar{y}_{t-1})^2 \right\}. \quad (60)$$

Here the difficulty comes from that we know neither T nor M . We use the iterative closet point (ICP) algorithm [9] to deal with this difficulty. The ICP algorithm is widely used for aligning two groups of points based on geometrical information. The ICP starts with a rough initial estimation on the transformation between the two groups of points, and then iteratively refines the transformation by identifying the matching points and minimizing an error metric. It can be proved that the ICP algorithm always converges to a local minimum when using the mean squared error objective function [9]. Different from classical ICP algorithm, here we do not need to generate matching points in each iteration since the candidate matching points are limited to the PS points. The details are described in Algorithm 1. In our experiments, the algorithm usually converges in a few iterations.

Algorithm 1 ICP based PS Matching Algorithm

- 1: **INPUT** Two groups of PSs: $G_{t-1} = \{PS_{t-1}^i\}$ and $G_t = \{PS_t^j\}$.
 - 2: **Initialization** Transformation parameters: $s = 1, \beta = 0, t_x = 0, t_y = 0$.
 - 3: **while** Matching M changes **do**
 - 4: Transform $\{PS_{t-1}^i\}$ into $\{PS_{t-1}^{i'}\}$ using (48) and the parameters (s, β, t_x, t_y) .
 - 5: For each $PS_{t-1}^{i'}$, find its nearest PS in G_t , denoted by PS_t^j , where

$$j = \arg \min_{k \in D(i)} \text{dis}(PS_{t-1}^{i'}, PS_t^k). \quad (61)$$
 - If $\text{dis}(PS_{t-1}^{i'}, PS_t^j) < d_{th}$, set $M(i) = j$; else, set $M(i) = 0$. (dis denotes the Euclidean distance between two PSs and d_{th} is a distance threshold to remove outliers.)
 - 6: Recalculate the parameters (s, β, t_x, t_y) by using current matching M and (50)–(53).
 - 7: **end while**
 - 5: **OUTPUT** Matching M and transformation $T = (s, \beta, t_x, t_y)$.
-

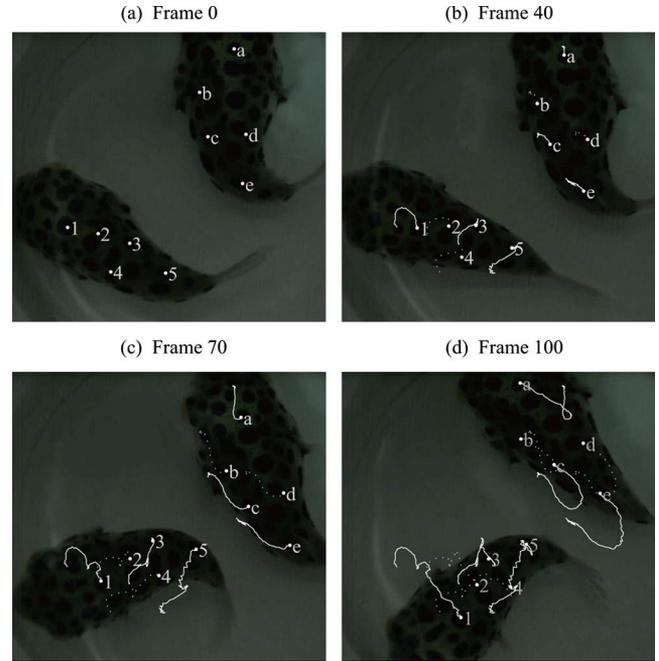


Fig. 7. Examples of PSs' trajectories. Solid lines and dash lines represent the positive and negative PS trajectories respectively. Small plates denote the PSs in current frame.

We apply the proposed method on the video of swimming fugu in water. The video is recorded as 30 frames/second and each frame is of size 512×480 . We calculate the PSs for each frame and the scale is fixed as 5. We track the corresponding PSs in the frames by the above method and record the trajectories of the PSs. The main computation here comes from the PS calculation, which takes 0.054 s per frame by a PC with an Intel 1.8G CPU. Several example trajectories with the frame images are shown in Fig. 7. More trajectories are available in a video which show the trajectories existing more than 90 frames (<http://www.gavo.t.u-tokyo.ac.jp/~qiao/fugu.avi>). The two fugu undertake clear shift and rotation motions in the video. The images also exhibit illumination changes due to waves. A number of PSs can be detected within the fugu bodies, which are separated from each other and cover the whole body of fugu. The trajectories of the PSs provide detailed information on the local motion of the fugu's body, which can be used for further biological kinematic analysis.

We also make comparison with the Harris corner (HC) [47] based tracking method. We use the implementation given by [48] to calculate Harris corners. To have a fair comparison, we set the threshold for determining HC as the one such that the number of Harris corners and the number of PSs are almost the same. We determine the corresponding relations of HCs in two continuous frames and calculate the trajectories of Harris corners with the same method for PS. For every two continuous frames, we calculate the successful matching-point (HC or PS) number (SMPN) and the successful matching rate (SMR) of SMPN to the total point number in the previous frame. In the experiment with Harris corners, the average SMPN is 253 and the average SMR is 58.9%;

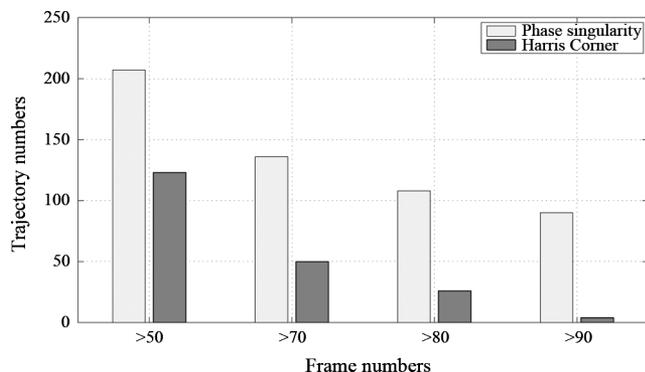


Fig. 8. Histogram of trajectory's length.

while in the PS-based experiment, the average SMPN is 302 and the average SMR is 73.0%. Our method can find more successful matching points with higher successful matching rates. In object tracking, one is also interested in the feature points (PS or HC) which can exist in a number of continuous frames. We estimate the number of PS or HC trajectories which exist in more than 50, 70, 80 and 90 frames. The results are depicted in Fig. 8. It can be seen that the PS-based method can find markedly more longer trajectories than the HC-based method. We also make a video of HC-based fugu tracking (http://www.gavo.t.u-tokyo.ac.jp/~qiao/fugu_HC.avi), which shows the trajectories with length more than 70 frames. There is only one trajectory found in the down-left fugu, which conducts a larger movement than the other fugu. On the other hand, there exist a number of trajectories on both the fugues in the PS-based experiment. This indicates that the PS-based method is more robust to the distortion caused by fugu motion.

Finally, it is not our objective to develop the best object tracking system in this paper. Here we only evaluate the usage of PSs for precise tracking without considering motion models. Moreover, it is possible to incorporate the PSs into a statistical tracking framework by using the probability model (30) and the scale stability analysis (38) and (39).

V. APPLICATION 2: IMAGE MATCHING

In this section, we describe another application of PSs: image matching. Using PSs, the image matching problem can be seen as identifying the corresponding pairs of PSs in two images. Different from the object tracking task in Section IV where two adjacent frames usually have similar scale and texture, we deal with relatively significant transformations and deformations between two matched images in this section. We have shown that PSs are invariant to translation and rotation (Section II) of an image, and the scaling problem can be further solved through PS selection in scale space (Section III-C). Thus, the image matching problem is reduced to how to match the selected key PSs.

The challenge here comes from the fact that natural images contain complex patterns and it is difficult to uniquely identify

corresponding PS pairs only using the local core measures presented in Section II-A; moreover, the two matched images do not necessarily have the same contents due to viewpoint and scale changes, and, thus, there may exist a number of PSs which do not have counterparts. Therefore, a rich and discriminant description of PSs is necessary for precious image matching. Generally, a rich description can be calculated as a feature vector for a local image patch associated with a PS point. Here we do not require image segmentation or object detection as preprocessing, since the size of the patch can be determined by the selected scale. The rich descriptions are also called local descriptors (features). Local descriptors have been successfully used for various tasks such as image matching, object recognition, and category classification. Mikolajczyk and Schmid [49] carried out an extensive comparison study of various local descriptors. Their results indicate that the scale invariant feature transform (SIFT) based descriptors proposed by Lowe [10] perform the best. For this reason, we adopt the SIFT descriptor for our application.

The classical SIFT approach [10] identifies a key point by comparing its difference-of-Gaussian value with those of its eight neighbors in 3-D scale space. Unlike this, here the key points with associated scales are determined by finding the extreme points of normalized derivatives along PS curves. For each PS point, we can crop a circled patch in the image centered at this point, whose radius is determined by the associated scale. Then we calculate the SIFT descriptor from this patch. The details are as follows. At first, the image patch is divided into 4×4 subregions. Secondly, we calculate the image gradient magnitudes and orientations for every pixel in a subregion. Then, the magnitudes are weighted by a 2-D Gaussian function centered at the PS point, and the gradient angles are quantized into eight orientations. Finally, we accumulate the weighted magnitudes for each quantized orientation to obtain a histogram representation in every subregion. The SIFT descriptor for the whole patch has a total dimensionality of $8 \times 4 \times 4 = 128$. SIFT descriptors are highly discriminant in the sense that it can be used to correctly identify two matching points among a larger number of candidates [10]. They are also robust to illumination changes, noise, and occlusion.

We evaluate the above image matching method on the benchmark database¹[33]. The database contains eight sets of structured and textured images. Each set includes a reference image and five compared images under different types of variations, such as viewpoint, scale, illumination, blur, and JPEG compression. The homographies (affine transformation matrixes) the reference image and the compared images in a particular set are available, which allows us to examine the correctness of the identified corresponding relations. For each image, we calculate the key PS points with extremal normalized Laplacian value and determine their associated scales. For a key PS p_1 in a reference image, we calculate its corresponding point p'_1 and scale s'_1 in a test image based on the homography matrix. If there exists a key PS p_2 with scale s_2 in the reference image where

¹Available at <http://www.robots.ox.ac.uk/~vgg/research/affine/index.html>

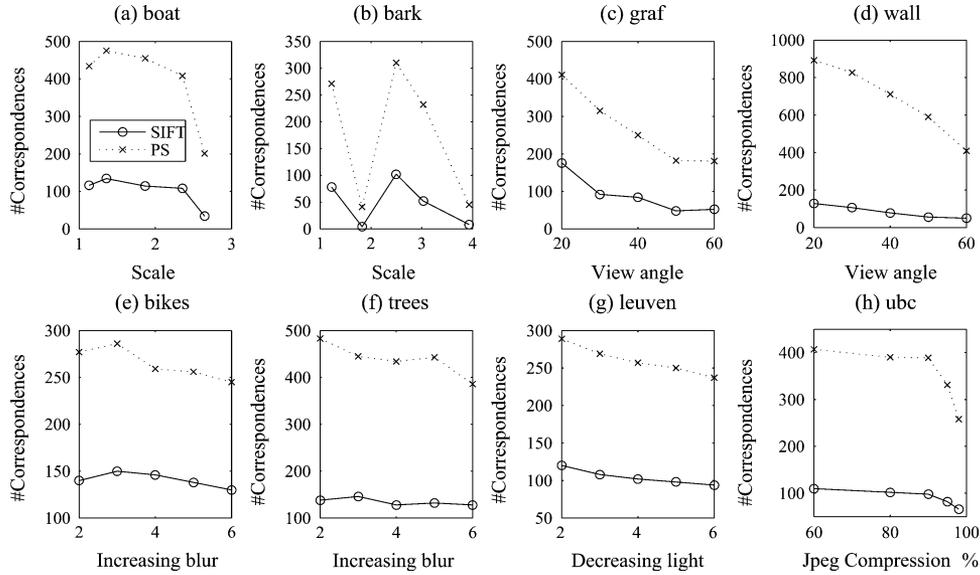


Fig. 9. Numbers of matching points under various variations.

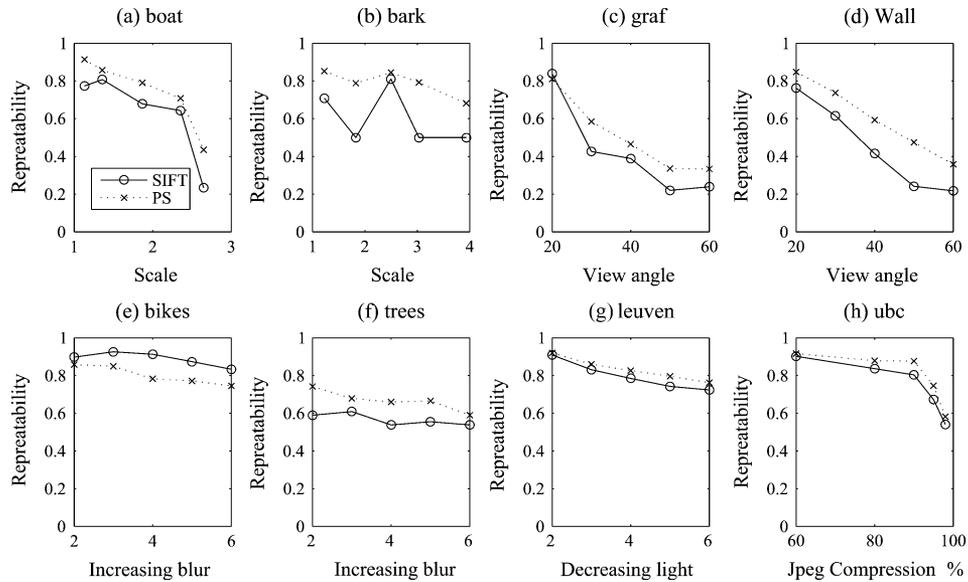


Fig. 10. Repeatability rates under various variations.

$|p'_1 - p_2| < 2 * s'_1$ and $s'_1/s_2 < 1.2$, we call p_1 and p'_1 a correct matching pair. We count the number of correct matching pairs among the key points in the two images, and calculate the repeatability rate of the key PS with correct matching pair in the reference image. The scale range for searching PS or SIFT points are set as [5], [10] in the reference images, and the scale range for the testing images are set according to the rough scale difference between the two images. We make comparisons with Lowe’s key point detection method [10], since both methods are scale invariant (not fully affine invariant as those in [33]). The parameters are set the same for both methods. The experimental results are summarized in Figs. 9 and 10.

For all the comparisons, the proposed method always can find more correct matching pairs than Lowe’s detector (Fig. 9). This

is a desirable fact that allows us to obtain more matching points. Moreover, except the “bikes” image set, the repeatability rates of our method are always higher than those of Lowe’s detector (Fig. 10). This indicates that the selection of key PSs are usually more robust than Lowe’s detector for the image matching task. Five examples of image matching based on key PS points are shown in Fig. 11. It can be seen that the proposed method can find a large number of correct matching points for each pair, which distribute over the whole images.

VI. CONCLUSION

This paper develops a theory of phase singularities for image representation. We show that PSs calculated by the Laguerre-Gauss filter contain important information of an image, and

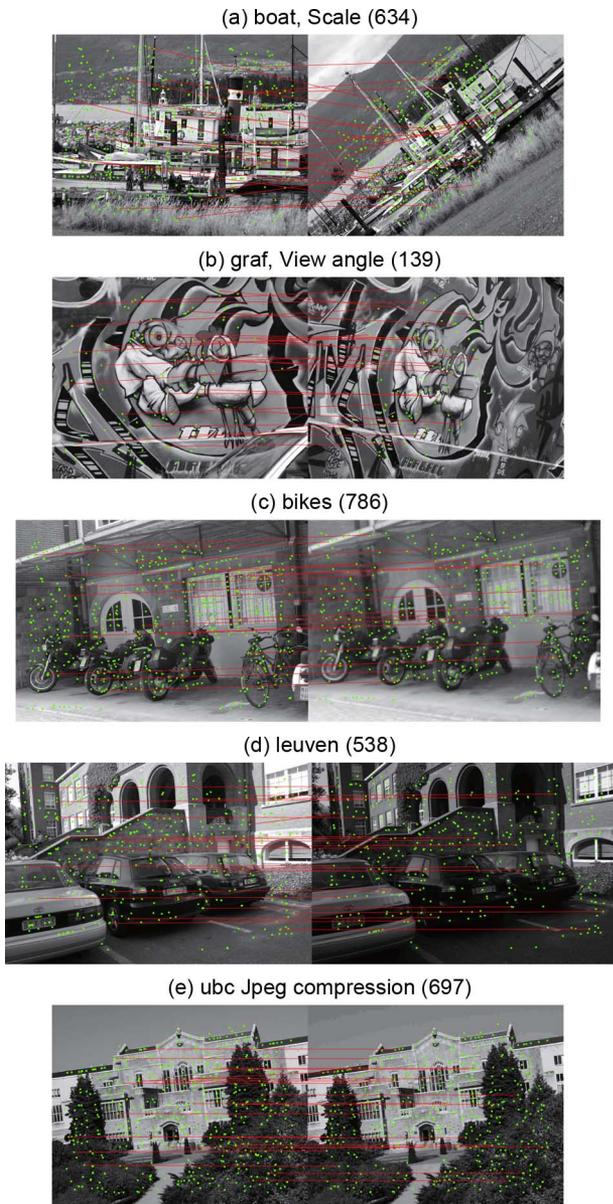


Fig. 11. Examples of image matching. The numbers in brackets denote the numbers of matching pairs found. Green points represent all the matching points. For easy observation, we only draw 20 matching pairs (red lines) in each example.

provide an efficient and effective tool for image analysis and presentation. PSs are invariant to translation and rotation. We analyze the local core structure about PSs, and discuss the relations among PSs, level-sets, and zero level-sets. Several invariant measures are introduced to characterize the core structure of PSs. The stability of PSs to noise addition is analyzed both theoretically and experimentally. We also study PSs in scale space where PS points connect to form PS curves. We show how to select key PS points with characteristic scales along PS curves by maximizing (minimizing) the normalized derivatives. The PS selection allows us to solve the image scaling problem.

To examine the usefulness of PSs, we develop two applications: object tracking and image matching. In the first application, we show that one can precisely locate the tracked objects by using the spatial structure of PSs. In the second application, we combine key PSs with the SIFT descriptor for robust image matching. The experimental results on a benchmark database [33] indicate that our method achieves comparable results with Lowe's well-known method [10], [33]. Although this paper is limited to PSs calculated by the LG filter, many results can be applied or generalized to other complex filters. The two promising applications have verified the usefulness of the proposed theory of PSs for image representation. We believe that there will be more applications of phase singularity in future work.

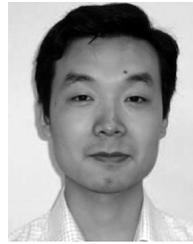
ACKNOWLEDGMENT

The authors would like to thank R. Ishijima and D. Honda for recording the fugu video.

REFERENCES

- [1] T. Huang, J. Burnett, and A. Deczky, "The importance of phase in image processing filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-23, no. 6, pp. 529–542, Jun. 1975.
- [2] D. J. Fleet *et al.*, "Phase-based disparity measurement," *CVGIU*, vol. 53, no. 2, pp. 198–210, 1991.
- [3] D. J. Fleet and A. D. Jepson, "Stability of phase information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 12, pp. 1253–1268, Dec. 1993.
- [4] G. Carneiro and A. D. Jepson, "Phase-based local features," in *Proc. ECCV*, 2002, pp. 282–296.
- [5] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE*, vol. 69, no. 5, pp. 529–541, May 1981.
- [6] J. L. Horner and P. D. Gianino, "Phase-only matched filtering," *Appl. Opt.*, vol. 23, no. 6, pp. 812–816, 1984.
- [7] Z. Xiaoxun and J. Yunde, "Local steerable phase (LSP) feature for face representation and recognition," in *Proc. CVPR*, 2006, vol. 2.
- [8] A. D. Jepson and D. J. Fleet, "Phase singularities in scale-space," *Image Vis. Comput.*, vol. 9, no. 5, pp. 338–343, 1991.
- [9] P. J. Besl and H. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
- [10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [11] Y. Qiao, W. Wang, N. Minematsu, J. Liu, and X. Tang, "Phase singularities for image representation and matching," in *Proc. ICASSP*, 2008.
- [12] J. F. Nye and M. V. Berry, "Dislocations in wave trains," *Proc. Roy. Soc. London A*, vol. 336, no. 1605, pp. 165–190, 1974.
- [13] L. Gil, "Vector order parameter for an unpolarized laser and its vectorial topological defects," *Phys. Rev. Lett.*, vol. 70, no. 2, pp. 162–165, 1993.
- [14] P. Couillet, L. Gil, and F. Rocca, "Optical vortices," *Opt. Commun.*, vol. 73, pp. 403–408, 1989.
- [15] M. Berry, "Making waves in physics. Three wave singularities from the miraculous 1830s," *Nature*, vol. 403, no. 6765, p. 21, 2000.
- [16] M. Berry and M. Dennis, "Phase singularities in isotropic random waves," *Roy. Soc. Proc.: Math., Phys. Eng. Sci.*, vol. 456, no. 2001, pp. 2059–2079.
- [17] W. Wang, T. Yokozeki, R. Ishijima, M. Takeda, and S. G. Hanson, "Optical vortex metrology based on the core structures of PS in LG transform of a speckle pattern," *Opt. Exp.*, vol. 14, no. 22, pp. 10195–10206, 2006.
- [18] W. Wang, Y. Qiao, R. Ishijima, T. Yokozeki, D. Honda, A. Matsuda, S. G. Hanson, and M. Takeda, "Constellation of phase singularities in the complex signal representation of a speckle-like pattern and its applications to optical vortex metrology for biological kinematic analysis," *Opt. Exp.*, vol. 16, no. 18, pp. 13908–13917, 2008.

- [19] C. S. Guo, Y. J. Han, J. B. Xu, and J. Ding, "Radial Hilbert transform with Laguerre-Gaussian spatial filters," *Opt. Lett.*, vol. 31, no. 10, pp. 1394–1396, 2006.
- [20] W. Wang, T. Yokozeki, R. Ishijima, A. Wada, Y. Miyamoto, S. G. Hanson, and M. Takeda, "Optical vortex metrology for nano metric speckle displacement measurement," *Opt. Exp.*, vol. 14, no. 1, pp. 120–127, 2006.
- [21] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.
- [22] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biol. Cybern.*, vol. 61, no. 2, pp. 103–113, 1989.
- [23] J. Daugman, "Statistical richness of visual phase information: Update on recognizing persons by IRIS patterns," *Int. J. Comput. Vis.*, vol. 45, no. 1, pp. 25–38, 2001.
- [24] J. B. Martens, "The Hermite transform-theory," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 9, pp. 1595–1606, Sep. 1990.
- [25] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 79–116, 1998.
- [26] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Jun. 1986.
- [27] R. A. Young, "The Gaussian derivative model for spatial vision: I. Retinal mechanisms," *Spatial Vis.*, vol. 2, no. 4, pp. 273–293, 1987.
- [28] M. Monastyrsky, "Topology of gauge fields and condensed matter," *Uspekhi Fizicheskikh Nauk*, vol. 166, no. 2, pp. 219–220, 1996.
- [29] I. Freund and N. Shvartsman, "Wave-field phase singularities: The sign principle," *Phys. Rev. A*, vol. 50, no. 6, pp. 5164–5172, 1994.
- [30] I. Freund, "Saddles, singularities, and extrema in random phase fields," *Phys. Rev. E*, vol. 52, no. 3, pp. 2348–2360, 1995.
- [31] F. Kanters, L. Florack, B. Platel, and B. M. ter Haar Romeny, "Image reconstruction from multiscale critical points," presented at the Scale Space Methods in Computer Vision: 4th Int. Conf., Scale-Space 2003, 2003.
- [32] M. Reeken, "Stability of critical points under small perturbations Part I: Topological theory," *Manuscripta Math.*, vol. 7, no. 4, pp. 387–411, 1972.
- [33] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1, pp. 43–72, 2005.
- [34] T. Iijima, "Basic theory on normalization of a pattern (in case of typical one-dimensional pattern)," *Bull. Elect. Lab.*, vol. 26, pp. 368–388, 1962.
- [35] A. P. Witkin, "Scale space filtering," in *Proc. IJCAI*, 1983, pp. 1019–1022.
- [36] J. J. Koenderink, "The structure of images," *Biol. Cybern.*, vol. 50, no. 5, pp. 363–370, 1984.
- [37] T. Lindeberg, *Scale-Space Theory in Computer Vision*. New York: Springer, 1994.
- [38] J. Damon, "Local Morse theory for solutions to the heat equation and Gaussian blurring," *J. Diff. Eq.*, vol. 115, no. 2, pp. 368–401, 1995.
- [39] T. Poston and I. Stewart, *Catastrophe Theory and Its Applications*. New York: Dover, 1996.
- [40] L. Florack and A. Kuijper, "The topological structure of scale-space images," *J. Math. Imag. Vis.*, vol. 12, no. 1, pp. 65–79, 2000.
- [41] R. Thom, *Structural Stability and Morphogenesis*. New York: Perseus, 1989.
- [42] P. Johansen, "On the classification of topoints in scale space," *J. Math. Imag. Vis.*, vol. 4, no. 1, pp. 57–67, 1994.
- [43] A. Kuijper, L. M. J. Florack, and M. A. Viergever, "Scale space hierarchy," *J. Math. Imag. Vis.*, vol. 18, no. 2, pp. 169–189, 2003.
- [44] A. Kuijper and L. M. J. Florack, "Hierarchical pre-segmentation without prior knowledge," in *Proc. ICCV*, 2001, pp. 487–493.
- [45] K. Mikolajczyk, "Detection of Local Features Invariant to Affine Transformations," Ph.D. dissertation, 2002.
- [46] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, 2006.
- [47] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vision Conf.*, 1988, vol. 15, p. 50.
- [48] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry, An Invitation to 3D Vision, from Images to Models 2003 [Online]. Available: <http://vision.ucla.edu/masks/>
- [49] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.



Yu Qiao (S'05-M'07) received the B.E. and M.E. degrees from Chongqing University, China, in 2000 and 2003, respectively, and the Ph.D. degree from the University of Electro-Communications, Japan, in 2006.

He was a postdoctoral researcher with the Chinese University of Hong Kong from 2006 to 2007. Then he was a JSPS fellow with the University of Tokyo. Now he is a researcher with the University of Tokyo. His research interests include speech recognition, image processing, computer

vision, pattern recognition, and machine learning.

Dr. Qiao is a member of ISCA and the Acoustical Society of Japan.



Wei Wang received the B.Sc. and D.Sc. degrees in physics from University of Science and Technology of China in 1997 and 2001, respectively, and the Ph.D. degree (summa cum laude) in optics from the University of Electro-Communications, Japan, in 2006.

He was a research fellow of the Japan Society for the Promotion of Science (JSPS) from 2003 to 2006 and has been a COE (Center of Excellence) postdoctoral research fellow since April 2006. He is now a Lecturer of optical engineering in the Department of

Mechanical Engineering, Heriot-Watt University, Edinburgh, U.K. His research interests include statistical optics, optical information processing, imaging science, and their applications to optical metrology.

Dr. Wang is a member of the SPIE, OSA, and JSAP.



Nobuaki Minematsu (M'06) received the Ph.D. degree in electronic engineering in 1995 from the University of Tokyo.

In 1995, he was an Assistant Researcher with the Department of Information and Computer Science, Toyohashi University of Technology, and in 2000, he was an Associate Professor with the Graduate School of Engineering, University of Tokyo. Since 2009, he has been an Associate Professor with the Graduate School of Information Science and Technology, University of Tokyo. From 2002 to 2003, he was a visiting researcher at Kungl Tekniska Högskolan (KTH), Sweden. He has a wide interest in speech from science to engineering, including phonetics, phonology, speech perception, speech analysis, speech recognition, speech synthesis, and speech applications.

Dr. Minematsu is a member of ISCA, IPA, the Computer Assisted Language Instruction Consortium, the Institute of Electronics, Information and Communication Engineering, the Acoustical Society of Japan, the Information Processing Society of Japan, the Japanese Society for Artificial Intelligence, and the Phonetic Society of Japan.



Jianzhuang Liu (M'02-SM'02) received the B.E. degree from the Nanjing University of Posts and Telecommunications, China, in 1983, the M.E. degree from Beijing University of Posts and Telecommunications, China, in 1987, and the Ph.D. degree from The Chinese University of Hong Kong in 1997.

From 1987 to 1994, he was a faculty member with the Department of Electronic Engineering, Xidian University, China. From August 1998 to August 2000, he was a research fellow at the School of

Mechanical and Production Engineering, Nanyang Technological University, Singapore. Then he was a postdoctoral fellow with The Chinese University of Hong Kong for several years. He is now an Assistant Professor in the Department of Information Engineering, The Chinese University of Hong Kong. His research interests include image processing, computer vision, pattern recognition, and machine learning.



Mitsuo Takeda received the B.S. degree in electrical engineering from the University of Electro-Communications (UEC) in 1969, and the M.S. and Ph.D. degrees in applied physics from the University of Tokyo, respectively, in 1971 and 1974.

After working for Canon, Inc., he joined the faculty of UEC in 1977. During 1985, he was a visiting scholar at the Information Systems Laboratory, Stanford University, Stanford, CA. He is currently a Professor in the Department of Information and Communication Engineering, UEC.

Dr. Takeda is Fellow of the three societies: SPIE (International Society for Optical Engineering), OSA (Optical Society of America), and JSAP (Japan Society of Applied Physics).



Xiaou Tang (S'93–M'96–SM'02–F'09) received the B.S. degree from the University of Science and Technology of China, Hefei, in 1990, the M.S. degree from the University of Rochester, Rochester, NY, in 1991, and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1996.

He is a Professor in the Department of Information Engineering, Chinese University of Hong Kong. He worked as the group manager of the Visual Computing Group at the Microsoft Research Asia from 2005 to 2008. His research interests include computer

vision, pattern recognition, and video processing.

Dr. Tang received the Best Paper Award at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2009. He is a program chair of the IEEE International Conference on Computer Vision (ICCV) 2009 and an Associate Editor of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and the *International Journal of Computer Vision*.