Technical Report DIKU-TR-97/18
Department of Computer Science
University of Copenhagen
Universitetsparken 1
2100 Copenhagen East
Denmark

August 1997
submitted for publication

Scale-Space has been Discovered in Japan

Joachim Weickert

Dept. of Computer Science, University of Copenhagen,
Universitetsparken 1, 2100 Copenhagen East, Denmark
E-mail: joachim@diku.dk

Seiji Ishikawa

Dept. of Control Engineering, Kyushu Institute of Technology,
Sensuicho 1-1, Tobata, Kitakyushu 804, Japan
E-mail: ishikawa@is.cntl.kyutech.ac.jp

Atsushi Imiya

Dept. of Information and Computer Sciences,
Faculty of Engineering, Chiba University,
1-33 Yayoi-Cho, Inage-Ku 263, Chiba, Japan
E-mail: imiya@ics.tj.chiba-u.ac.jp

Abstract

Gaussian scale-space is considered to be a modern bottom-up tool in computer vision. The American and European vision community, however, is unaware of the fact that Gaussian scale-space has already been axiomatically derived in 1959 in a Japanese paper by Taizo Iijima. This result formed the starting point of an entire world of linear scale-space research in Japan ranging from various axiomatic derivations over deep structure analysis to applications to optical character recognition (OCR). Since this world is unknown to western scale-space researchers and many papers are written in Japanese, we give an overview of the basic concepts. In particular, we review four Japanese axiomatics for Gaussian scale-space which have been proposed between 1959 and 1981. By juxtaposing them to ten American or European axiomatics, we present an overview of the state-of-the-art in Gaussian scale-space axiomatics.

Key words: Scale-space, axiomatics, deep structure, OCR.

1 Introduction

A rapidly increasing number of publications, workshops and conferences which are devoted to scale-space ideas confirms the impression that the scale-space paradigm belongs to the challenging new topics in computer vision.

In scale-space theory one embeds an image $f : \mathbb{R}^2 \to \mathbb{R}$ into a continuous family $\{T_t f \mid t \geq 0\}$ of gradually smoother versions of it. The original image corresponds to the scale $t = 0$ and increasing the scale should simplify the image without creating spurious structures. Since a scale-space introduces a hierarchy of the image features, it constitutes an important step from a pixel-related image description to a semantical image description.

Usually a 1983 paper by Witkin [67] or an unpublished 1980 report by Stansfield [60] are regarded as the first references to the scale-space idea. Witkin obtained a scale-space representation by convolution of the original image with Gaussians of increasing width. Koenderink [41] pointed out that this *Gaussian scale-space* is equivalent to calculating $(T_t f)(x)$ as the solution $u(x,t)$ of the linear diffusion process

$$\partial_t u \;=\; \sum_i \partial_{x_i x_i} u \;=:\; \Delta u, \qquad (1)$$

$$u(x,0) \;=\; f(x). \qquad (2)$$

Soon linear diffusion filtering became very popular in image processing, and many results have been obtained with respect to axiomatization, differential geometry, deep structure, and applications. An excellent overview of all these aspects can be found in the recent book edited by Sporring, Nielsen, Florack and Johansen [59].

Perona and Malik [53] pioneered the field of nonlinear diffusion processes, where the diffusivity is adapted to the underlying image structure. Many regularized variants of the Perona–Malik filter are well-posed and reveal scale-space properties [63, 64, 65]. Other important classes of nonlinear scale-space have been established as well. Some of them are continuous-scale versions of classical morphological processes such as dilation or erosion [4, 5, 38], others can be described as intrinsic evolutions of level curves [1, 40, 50, 56]. These scale-spaces are generated by nonlinear partial differential equations (PDEs) which are designed to have affine [1, 56] or projective invariances [11, 6, 50, 10]. Overviews of nonlinear approaches can be found in [9, 16, 64].

This diversity of scale-space approaches has triggered people to investigate which of these equations can be distinguished in a unique way from others, because they can be derived from first principles (axioms) [41, 68, 3, 44, 13, 4, 1, 50, 52, 48, 45, 12]. Apart from a few exceptions [4, 1, 50], all of these axiomatics use (explicitly or implicitly) one requirement: a linearity assumption. Within such a linear framework it was always possible to derive the Gaussian scale-space as the unique possibility. The fact that many of these approaches have been found recently shows that linear scale-space axiomatic belongs to the current research topics in computer vision.

However, since the linear diffusion equation is well-established in mathematics and physics since Fourier's pioneering work in 1822 [14], and image processing was already an active field in the fifties, one might wonder whether the concept of Gaussian scale-space is not much older as well. Koenderink [42] states in a very nice preface discussing how scale-related ideas can be traced back in literature, poetry, painting and cartography that *"the key ideas have been around for centuries and essentially everything important was around by the end of the nineteenth century."* The goal of the present paper is to supplement these statements by showing that not only the ingredients, but also their final mixture and application to image processing is much older than generally assumed. To this end we present four Japanese scale-space approaches, which are older than American and European ones. The first one of them dates back to 1959.

The outline of this paper is as follows: In Section 2 we describe the basic ideas of a 1-D axiomatic for Gaussian scale-space that has been discovered by Taizo Iijima in 1959 [20]. Section 3 studies a 2-D version of this axiomatics leading to affine Gaussian scale-space. It has been established in 1962. In 1971 Iijima presented a more physical 2-D axiomatics of affine Gaussian scale-space [25]. Its principles are sketched in Section 4. Section 5 describes a 2-D axiomatic which has been found by Nobuyuki Otsu in 1981 [51]. In Section 6 we shall relate all these results to the well-known linear axiomatics that have been established since 1984. We conclude with a discussion in Section 7. In order to give the reader an impression of the spirit of these Japanese papers, we stick closely to the notations in the original work. The discussions are supplemented with remarks on the philosophical background,

physical and biological motivations, results on the deep structure in scale-space and applications to optical character recognition (OCR).
A preliminary version of this paper focusing exclusively on the axiomatic aspects of 2 of these 4 Japanese frameworks has been presented in [66].

2 Iijima's 1-D axiomatic (1959)

2.1 Historical and philosophical background

Japanese scale-space research was initiated by Taizo Iijima. After graduating in electrical engineering and mathematics from Tokyo Institute of Technology in 1948, he joined the Electrotechnical Laboratory (ETL). In his thesis titled *'A fundamental study on electromagnetic radiation'* he derived the third analytical solution of the radiation equation. During these studies he acquired the mathematical and physical prereqisites for his later scale-space work.
At the ETL Iijima was involved in different research activities on speech and pattern recognition. Triggered by actual needs such as optical character recognition (OCR), but also voice typewriting, or medical diagnosis, he wanted to establish a general theoretical framework for extracting characteristic information of patterns. This framework should avoid extreme standpoints such as purely deterministic or purely stochastic classifications, and it should make use of the original physical or geometric characteristics of the patterns [24].
Besides this problem-driven background, there was also a philosophical motivation for Iijima's scale-space research. Its principles go back to Zen Buddhism, and they can be characterized by the sentence *"Anything is nothing, nothing is anything"*. Applied to the scale-space context this means to obtain the desired information, it is necessary to control the unwanted information. The blurred scale-space evolution may be interpreted as a kind of unwanted information which helps to understand the semantical content of the unblurred image. In this sense important information is existing in seemingly unimportant information.

2.2 Axioms

Iijima's first axiomatic formulation of the scale-space concept can be found in a technical paper from 1959 titled *'Basic theory of pattern observation'* [20]. A journal version of this paper has been published in 1962 under the title *'Basic theory on normalization of pattern (In case of typical one-dimensional pattern)'* [21]. Both papers are written in Japanese. The restriction to the 1-D case is for simplicity reasons. Extensions to 2-D are discussed in Section 3.
Iijima imposes basic principles which are in accordance with requirements from observation theory: a robust object recognition should be invariant under changes in the reflected light intensity, parallel shifts in position, and expansions or contractions of the object.
In addition to these three transformations he considers an observation transformation Φ which depends on an observation parameter σ and which transforms the original image $g(x)$ into a blurred version $f(x)$. This class of blurring transformations is called *'boke'* (defocusing). He assumes that it has the structure[1]

$$f(x) = \Phi[g(x'), x, \sigma] = \int_{-\infty}^{\infty} \phi\{g(x'), x, x', \sigma\}\, dx', \tag{3}$$

and that it should satisfy four conditions:

(I) *Linearity (with respect to multiplications):*
If the intensity of a pattern becomes A times its original intensity, then the same should happen to the observed pattern:

$$\Phi[Ag(x'), x, \sigma] \;=\; A\,\Phi[g(x'), x, \sigma]. \tag{4}$$

(II) *Translation invariance:*
Filtering a translated image is the same as translating the filtered image:

$$\Phi[g(x'-a), x, \sigma] \;=\; \Phi[g(x'), x-a, \sigma]. \tag{5}$$

[1] The variable x' serves as a dummy variable.

(III) *Scale invariance:*
If a pattern is spatially enlarged by some factor λ, then there exists a $\sigma' = \sigma'(\sigma, \lambda)$ such that

$$\Phi[g(x'/\lambda), x, \sigma] \;=\; \Phi[g(x'), x/\lambda, \sigma']. \tag{6}$$

(IV) *(Generalized) semigroup property:*
If g is observed under a parameter σ_1 and this observation is observed under a parameter σ_2, then this is equivalent to observing g under a suitable parameter σ_3:

$$\Phi\Big[\Phi[g(x''), x', \sigma_1], x, \sigma_2\Big] \;=\; \Phi[g(x''), x, \sigma_3], \tag{7}$$

where $\sigma_3 = \sigma_3(\sigma_1, \sigma_2)$, but not necessarily $\sigma_3 = \sigma_1 + \sigma_2$.

We recognize that the first three principles reflect just the requirements from observation theory. Later on we shall see that – in order to determine the Gaussian uniquely – this axiomatic has to be supplemented with a fifth requirement: preservation of positivity.

2.3 Consequences

In order to determine the function Φ Iijima establishes in a very systematic way four lemmas which start with the class (3) and confine this family by subsequently imposing one more of the conditions (I)–(IV):

(a) **Lemma 1:**
If Φ has the structure (3) and satisfies the linearity axiom (I), then it can be written as the integral

$$\Phi[g(x'), x, \sigma] \;=\; \int_{-\infty}^{\infty} g(x')\, \phi(x, x', \sigma)\, dx'. \tag{8}$$

(b) **Lemma 2:**
If Φ is given by (8) and satisfies the translation invariance axiom (II), then it can be written as a convolution operation:

$$\Phi[g(x'), x, \sigma] \;=\; \int_{-\infty}^{\infty} g(x')\, \phi(x - x', \sigma)\, dx'. \tag{9}$$

(c) **Lemma 3:**
If Φ is given by (9) and satisfies the scale invariance axiom (III), then it can be written as

$$\Phi[g(x'), x, \sigma] \;=\; \int_{-\infty}^{\infty} g(x')\, \phi(\nu(\sigma)(x - x'))\, \nu(\sigma)\, dx', \tag{10}$$

where $\nu(\sigma)$ is an arbitrary function of σ.

(d) **Lemma 4:**
If Φ is given by (10) and satisfies the semigroup axiom (IV), then it can be written either as

$$\Phi[g(x'), x, \sigma] \;=\; \int_{-\infty}^{\infty} g(x')\, \phi(\nu(\sigma)(x - x'))\, \nu(\sigma)\, dx' \tag{11}$$

where ϕ has the specific structure

$$\phi(u) \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-k^{2m} \xi^{2m} + i\xi u\right) d\xi \qquad (k \in \mathbb{R},\ m = 1, 2, ...), \tag{12}$$

or

$$\Phi[g(x'), x, \sigma] \equiv 0. \tag{13}$$

The proofs of the first three lemmas are rather short and not very complicated, whereas the longer proof of Lemma 4 involves some more sophisticated reasonings in the Fourier domain.
In a next step Iijima simplifies the result of Lemma 4. The case $\Phi[g(x'), x, \sigma] \equiv 0$ is of no scientific interest and is not considered any further. In equation (11) the function ν is eliminated by defining (without loss of generality) a new scale parameter σ via $\nu(\sigma) = k/\sigma$. Then the k-dependence in (11) and (12) immediately vanishes by means of substitution of variables. The results are summarized in a theorem, which states that Φ satisfying (3) and the axioms (I)–(IV) is given by

$$\Phi[g(x'), x, \sigma] \;=\; \int_{-\infty}^{\infty} g(x')\, \phi_m\left(\frac{x-x'}{\sigma}\right) \frac{dx'}{\sigma} \tag{14}$$

with

$$\phi_m(u) \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\xi^{2m} + i\xi u\right) d\xi \qquad (m = 1, 2, ...). \tag{15}$$

For this family it follows that σ' in (III) becomes $\sigma' = \sigma/\lambda$, and σ_3 in (IV) satisfies

$$\sigma_3^{2m} = \sigma_1^{2m} + \sigma_2^{2m}. \tag{16}$$

For the special case $m = 1$ equation (15) becomes

$$\phi_1(u) \;=\; \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{u^2}{4}\right), \tag{17}$$

which gives

$$\Phi[g(x'), x, \sigma] \;=\; \frac{1}{2\sqrt{\pi}\sigma} \int_{-\infty}^{\infty} g(x') \exp\left(\frac{-(x-x')^2}{4\sigma^2}\right) dx'. \tag{18}$$

Thus, $\Phi[g(x'), x, \sigma]$ is just the convolution between g and a Gaussian with standard deviation $\sigma\sqrt{2}$.
In another theorem he establishes that, if one requires that Φ is *positivity preserving*, i.e.

$$\Phi[f(x'), x, \sigma] > 0 \qquad \forall\, f(x) > 0, \quad \forall\, \sigma > 0, \tag{19}$$

then $m = 1$ arises by necessity. The proof presents an explicit example, where the positivity is not preserved for $m > 1$.
This concludes his axiomatic derivation of the Gaussian kernel under five assumptions: linearity, translation invariance, scale invariance, semi-group property, and preservation of positivity.
Interestingly, Iijima used his axiomatic also for justifying why humans and many animals have a visual system that is based on a lens [24]: an optical lens has a Gaussian-like blurring profile. In this sense, its existence can be regarded as a natural consequence from elementary observational principals. Conversely, it indicates that it is natural to require preservation of positivity.
Iijima considered scale-space as a first part of his theory of pattern recognition. In [24] one can find an overview of his ideas, the main 1-D results, their motivation from a viewpoint of observation theory and their theoretical foundation as a classification tool for OCR and other problems. This paper is written in English.

3 Iijima's 2-D axiomatic (1962)

Having obtained these one-dimensional results, it was straightforward for Iijima to generalize them to a two-dimensional scale-space axiomatic. This was done in a technical paper from 1962 [22], followed by a journal paper in 1963 [23], which also contains an interesting English abstract.
In this paper he considers a blurring transformation of type

$$f^*(x) \;=\; \Phi[f(x'), x, \Sigma] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi\{f(x'), x, x', \Sigma\}\, dx_1' dx_2', \tag{20}$$

where Σ is a positive definite 2×2 matrix. This transformation should satisfy four conditions:

(I) *Linearity (with respect to multiplications):*

$$\Phi[Ff(x'), x, \Sigma] \;=\; F\,\Phi[f(x'), x, \Sigma] \qquad \forall\, F \in \mathbb{R}. \tag{21}$$

(II) *Translation invariance:*

$$\Phi[f(x'-a), x, \Sigma] \;=\; \Phi[f(x'), x-a, \Sigma] \qquad \forall\, a \in \mathbb{R}. \tag{22}$$

(III) *Scale invariance and closedness under affine transformations:*
If the pattern is transformed by an (invertible) matrix Λ, then there exists a $\Sigma' = \Sigma'(\Sigma, \Lambda)$ such that

$$\Phi[f(\Lambda^{-1}x'), x, \Sigma] \;=\; \Phi[f(x'), \Lambda^{-1}x, \Sigma']. \tag{23}$$

(IV) *(Generalized) semigroup property:*
For every Σ_1 and Σ_2 there exists a $\Sigma_3 = \Sigma_3(\Sigma_1, \Sigma_2)$ such that

$$\Phi\Big[\Phi[f(x''), x', \Sigma_1], x, \Sigma_2\Big] \;=\; \Phi[f(x''), x, \Sigma_3]. \tag{24}$$

These four axioms in combination with the requirement of positivity preservation are sufficient to derive that the blurring transformation Φ is given by the affine Gaussian scale-space:

$$\Phi[f(x'), x, \Sigma] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1', x_2')\; \phi_1\Big(x_1 - x_1', x_2 - x_2', \Sigma\Big)\; dx_1'\, dx_2' \tag{25}$$

with

$$\phi_1(u_1, u_2, \Sigma) \;=\; \frac{1}{4\pi\sigma^2}\, \exp\left(-\frac{\mu_{22}u_1^2 - 2\mu_{12}u_1u_2 + \mu_{11}u_2^2}{4\sigma^2}\right) \tag{26}$$

and

$$\Sigma \;=\; \sigma^2 \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{12} & \mu_{22} \end{pmatrix}, \qquad \det \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{12} & \mu_{22} \end{pmatrix} \;=\; 1. \tag{27}$$

In order to single out the usual (isotropic) Gaussian scale-space, one has to impose one more axiom, namely invariance under rotations. Then axiom (III) should be reduced to pure scale invariance.

4 Iijima's 2-D axiomatic based on physical principles (1971)

4.1 The principles

In 1971 Iijima decided to reconsider his scale-space and pattern recognition theory in order to get to a consistent reformulation which leads to a simplification of his ideas.
As a result, 2-D affine Gaussian scale-space has been derived from physical principles. This is treated in a paper, which is available as a complete English translation in [25].
The goal of this paper is to obtain a generalized figure $f(r, \tau)$ from an original figure $f(r)$ in a way which is comparable with the defocusing of an optical system. Two principles are assumed to hold:

(I) *Conservation principle:*
The blurring transformation does not change the total light energy within the image. This means that the image satisfies the continuity equation

$$\frac{\partial f(r, \tau)}{\partial \tau} + \operatorname{div} I(r, \tau) = 0 \tag{28}$$

where the flux I denotes the figure flow.

(II) *Principle of maximum loss of figure impression:*
The figure flow is determined such that its normalized expression

$$J(I) \;:=\; \frac{\|I\, \nabla f\|^2}{I R^{-1} I} \tag{29}$$

takes the maximum value. Here, $R(\tau)$ denotes a positive definite matrix which is the medium constant of the blurring process.

From the last principle he derives that

$$I(r,\tau) \;=\; -R(\tau)\cdot\nabla f(r,\tau). \tag{30}$$

Together with the conservation principle this leads to the anisotropic linear diffusion equation

$$\frac{\partial f(r,\tau)}{\partial \tau} = \operatorname{div}\Big(R(\tau)\cdot\nabla f(r,\tau)\Big) \tag{31}$$

which is just the formulation of affine Gaussian scale-space as a partial differential equation. $R(\tau)$ is a diffusion tensor. Iijima calls this equation the *basic equation of figure.*

4.2 Further results by Iijima and his students

In the remainder of [25] Iijima shows that this equation is essentially invariant under conformal transformations describing multiplication of grey values with a constant, addition of linear brightness gradients, translations, and affine transformations. Essentially invariant means that it may be transformed into another anisotropic linear diffusion equation.

The set of these conformal mappings form an algebraic group, while the set of affine Gaussian blurring transformations form a semigroup. Iijima studies compositions of a conformal mapping and a blurring transformation under the name observational transformations. These transformations are used to construct a theory of pattern classification: two figures are considered as equivalent if they result from the same original figure by observational transformations.

Iijima compares the invariance of the basic equation of figure under conformal transformations with the invariance of Newton's basic equations of motion to Galileo's transformation, and the invariance of the Maxwell equations to the Lorentz transformation. It seems that he was fully aware of the future importance of his discovery when he wrote in [25] that *"this paper provides a basis for exploring the recognition theory of visual patterns and solving mathematically the various problems in visual physiology".*

This subsequent recognition theory is documented in a series of Japanese papers which are either available as full English translations [26, 27, 28, 29, 30] or as extended English abstracts [31, 32]. Iijima regards a Gaussian-blurred figure as an element in a Hilbert space which can be expanded in an orthonormal function system given in terms of Hermite polynomials. The similarity between two patterns is a function of the scalar product in this Hilbert space. Gaussian blurring plays a central role in this theory, because it makes the algorithms insensitive to noise, it reduces effects of positional deviation, and it allows a coarser sampling. In order to overcome the problem that blurred patterns become more similar, he devised a specific canonical transformation. Incorporating all these features and modifying the similarity measure has lead Iijima to a robust scale-space based pattern matching technique which he called *multiple similarity method.*

Applications of Iijima's theory to OCR have been presented in English proceedings papers at the First USA–Japan Computer Conference in 1972 [35], and at the First International Joint Conference on Pattern Recognition in 1973 [36]. In [35] it is described how Iijima, Genchi and Mori have realized the multiple similarity method in hardware in the optical character reader ASPET/71. This machine was capable of reading 2000 alphanumeric characters per second, and the scale-space part has been regarded as the reason for its reliability and robustness. Others stated about ASPET/71 that *"it has been proved to have better performance than any similar conventional method"* [49]. It is now exhibited at the National Museum of Sciences in Tokyo. The ASPET/71 was an analog machine, but its commercial variant OCR-V100 by Toshiba used digital technology fully. Iijima's multiple similarity method has also become the main algorithm of Toshiba's later OCR systems [47].

In 1973 Iijima condensed his whole scale-space and pattern classification theory to a Japanese textbook [33]. It can be regarded as one of the first monographs on linear scale-space theory.

In the eighties he addressed together with Nan-yuan Zhao, a Ph.D. student of him, the problem of deep structure analysis in scale-space [70]; see also the discussion in [37]. For a solution $f(x,t)$ of the isotropic linear diffusion equation, they constructed a curve which comprises the stationary points, i.e. locations (x,t) with $\nabla f(x,t) = 0$. This so-called *stationary curve* $r(t)$ obeys the equation

$$\operatorname{Hess}(f)\,\frac{dr(t)}{dt} = -\Delta\Big(\nabla f(r,t)\Big). \tag{32}$$

They stated criteria for identifying *stable viewpoints* on the stationary curve, for instance by requiring that $\frac{dr(t)}{dt}$ vanishes there without vanishing in a neighbourhood. Afterwards they linked these stable viewpoints to a topological scale-space tree [71]. It provides an hierarchical organization of extrema and saddle points at stable scales. To each stable point (x_i, t_i) they assign a region of interest which is given by a disk with center x_i and radius t_i. Applied to an image of Zhao himself, this focus-of-attention method extracted eyes, nostrils and the mouth as regions of interest [69].
Parts of this work on scale-space trees was further pursued by the image processing group of Makoto Sato, another former Ph.D. student of Iijima. Sato's group established linear scale-space results ranging from deep structure analysis [57, 58, 62, 7] to the filtering of periodic or spherical patterns [61, 39]. All cited Sato papers are written in English.
Iijima continued his research on scale-space techniques for OCR till the nineties [2]. After 1972 he held professorships at Tokyo Institute of Technology, Tokyo Engineering University, and the Advanced Institute of Science and Technology. In spring 1997 he retired at the age of 72. An English bibliography can be found in [2].

5 Otsu's 2-D axiomatic (1981)

5.1 Derivation of the Gaussian

In 1981 another Japanese scale-space axiomatic has been established in the Ph.D. thesis of Nobuyuki Otsu [51]. He wrote his thesis at the ETL, where Iijima was working in the sixties. Otsu derives two-dimensional Gaussian scale-space in an axiomatic way by modifying the axioms described in Section 2. Section 4.1 of his thesis is titled *'Axiomatic derivation of the scale transformation'*. There he considers some transformation of an image f into an image $\tilde{f}$, for which the following holds:

(I) *Representation as a linear integral operator:*
There exists a function $W : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ such that

$$\tilde{f}(r) \;=\; \int\limits_{\mathbb{R}^2} W(r, r')\, f(r')\, dr' \qquad \forall\, r \in \mathbb{R}^2. \tag{33}$$

(II) *Translation invariance:*
For all $r \in \mathbb{R}^2$ and for all $a \in \mathbb{R}^2$ it is required that

$$\tilde{f}(r-a) \;=\; \int\limits_{\mathbb{R}^2} W(r, r')\, f(r'-a)\, dr'. \tag{34}$$

Since this is just $\int_{\mathbb{R}^2} W(r, r'+a)\, f(r')\, dr'$, and (I) states that $\tilde{f}(r-a) = \int_{\mathbb{R}^2} W(r-a, r')\, f(r')\, dr'$, it follows that the integral kernel is symmetric,

$$W(r, r'+a) \;=\; W(r-a, r'), \tag{35}$$

and, thus, it is a convolution kernel:

$$W(r, r') \;=\; W(r - r'). \tag{36}$$

(III) *Rotation invariance (of the kernel):*
For all rotation matrices T_Θ and for all $r = (x, y)^T \in \mathbb{R}^2$ it is assumed that

$$W(T_\Theta r) \;=\; W(r). \tag{37}$$

Hence, W depends only on $|r|$: $W(r) = W(x^2 + y^2)$.

(IV) *Separability:*
There exists a function $u : \mathbb{R} \to \mathbb{R}$ such that

$$W(r) \;=\; u(x)\, u(y). \tag{38}$$

Combining this with (III) implies after elementary manipulations that

$$W(r) \;=\; k\,\exp\left[c\,(x^2+y^2)\right]$$

with some parameters $k, c \in \mathbb{R}$. In order to get $k > 0$ and $c < 0$, however, additional constraints are needed.

(V) His next requirement which he names "Normalization of energy" actually consists of two parts: *Preservation of nonnegativity,*

$$\tilde{f}(r) \geq 0 \qquad \forall\, f(r) \geq 0, \tag{39}$$

and *average grey level invariance,*

$$\int_{\mathbb{R}^2} \tilde{f}(r)\,dr \;=\; \int_{\mathbb{R}^2} f(r)\,dr. \tag{40}$$

This leads to $W(r) \geq 0$ and $\int_{\mathbb{R}^2} W(r)\,dr = 1$, respectively.

Combining these results gives $k = \frac{1}{2\pi\sigma^2}$ and $c = -\frac{1}{2\sigma^2}$. This yields the Gaussian kernel

$$W(r) \;=\; \frac{1}{2\pi\sigma^2}\,\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{41}$$

and concludes the axiomatic derivation of the 2-D linear scale-space.

5.2 Further results

Section 4.2 of Otsu's thesis is titled *'Representation of scale-space transformation and semigroup'*. It is devoted to the N-dimensional Gaussian scale-space. With $\rho := \sigma^2/2$ he defines

$$T(\rho)f(r) := \frac{1}{(4\pi\rho)^{N/2}}\,\exp\left(-\frac{|r|^2}{4\rho}\right) * f(r). \tag{42}$$

Using Fourier techniques he shows that the generator of the scale-space transformation is the Laplacean:

$$\tilde{f}(r,\rho) \;=\; T(\rho)f(r) \;=\; \exp(\rho\Delta)\,f(r). \tag{43}$$

This gives

$$\frac{\partial \tilde{f}(r,\rho)}{\partial \rho} \;=\; \Delta\left(\exp(\rho\Delta)f(r)\right) \;=\; \Delta\,\tilde{f}(r,\rho).$$

Thus, $\tilde{f}$ satisfies the isotropic linear diffusion equation.
The formal inversion of the scale-space transformation by means of (43) is

$$f(r) \;=\; [T(\rho)]^{-1}\tilde{f}(r,\rho) \;=\; \exp(-\rho\Delta)\,\tilde{f}(r,\rho) \;=\; \left(I - \rho\Delta + \frac{\rho^2}{2}\Delta^2 - ...\right)\tilde{f}(r,\rho).$$

For the case that ρ or $\Delta^2\tilde{f}$ is small, Otsu proposes to approximate $[T(\rho)]^{-1}$ by $[I - \rho\Delta]$ and to use it for recovering the original image from a blurred one[2].

6 Relation to other work

Having sketched the basic ideas of these Japanese axiomatics, it is natural to ask about similarities and differences to other approaches. Table 1 gives an overview of the current axiomatics for the continuous Gaussian scale-space. These axioms and some of their relations can be explained as follows[3]:

[2]This is an ill-posed problem which may lead to unstable results.

[3]Of course, such a table can only give a "flavour" of the different approaches, and the precise description of each axiom may slightly vary from paper to paper. Several relations between the presented axioms are discussed in [1, 45, 52].

Table 1: Overview of continuous Gaussian scale-space axiomatics (I1 = Iijima [20, 21], I2 = Iijima [22, 23], I3 = Iijima [25], O = Otsu [51], K = Koenderink [41], Y = Yuille/Poggio [68], B = Babaud et al. [3], L1 = Lindeberg [44], F1 = Florack et al. [13], A = Alvarez et al. [1], P = Pauwels et al. [52], N = Nielsen et al. [48], L2 = Lindeberg [45], F2 = Florack [12]).

	I1	I2	I3	O	K	Y	B	L1	F1	A	P	N	L2	F2
convolution kernel	•	•		•		•	•	•	•	•	•		•	•
semigroup property	•	•						•	•	•	•	•	•	•
locality										•				
regularity						•	•	•	•	•	•		•	•
infinetesimal generator											•			
max. loss principle			•											
causality					•	•	•	•					•	
nonnegativity	•	•		•						•	•			•
Tikhonov regularization												•		
average grey level invar.			•	•			•	•		•	•			
flat kernel for $t \to \infty$						•			•					
isometry invariance		•		•		•	•	•	•	•	•	•	•	•
homogeneity & isotropy					•									
separability				•					•					
scale invariance	•	•				•	•		•		•	•		•
valid for dimension	1	2	2	2	1,2	1,2	1	1	> 1	N	1,2	N	N	N

- **Convolution kernel:**
  There exists a family of functions $\{k_t : \mathbb{R} \to \mathbb{R} \mid t \geq 0\}$ such that

$$(T_t f)(x) = \int_{\mathbb{R}^N} k_t(x - x')\, f(x')\, dx'.$$

  In Section 5.1 we have already seen that this property can be derived from the two assumptions:

  - **Linear integral operator:**
    There exists a family of kernels $\{k_t \mid t \geq 0\}$ with

$$(T_t f)(x) = \int_{\mathbb{R}^N} k_t(x, x')\, f(x')\, dx'.$$

    Since every continuous linear functional can be written as an integral operator, it follows that Florack's topological duality paradigm [12] can also be interpreted as requiring the existence of a linear integral operator[4].

  - **Translation invariance:**
    Let a translation τ_a be defined by $(\tau_a f)(x) := f(x-a)$. Then,

$$\tau_a T_t = T_t \tau_a \qquad \forall\, a \in \mathbb{R}^N, \quad \forall\, t > 0.$$

  Since usually linearity and translation invariance are imposed in conjunction, we have summarized them under the term "convolution kernel".

- **Semigroup property:**

$$T_{t+s} f = T_t(T_s f) \qquad \forall\, t, s \geq 0, \quad \forall\, f.$$

  This property ensures that one can implement the scale-space process as a cascade smoothing which resembles certain processes of the human visual system.

[4]Dirac point distributions and their derivatives are admitted as "functions under the integral".

- **Locality:**
  For small t the value of $T_t f$ at any point x is determined by its vicinity:

  $$\lim_{t \to 0^+} (T_t f - T_t g)(x) = \mathrm{o}(x)$$

  for all $f, g \in \mathrm{C}^\infty$ whose derivatives of order ≥ 0 are identical.

- **Regularity:**
  A precise definition of the smoothness requirements for the scale-space operator depends on the author:

  - Since the original image creates the scale-space, it is natural to assume that it is continuously embedded, i.e. $\lim_{t \to 0^+} T_t = I$. In the linear convolution case, this means that $k_t(x)$ tends to Dirac's delta distribution [68] and its Fourier transform becomes 1 everywhere [13].
  - Babaud et al. [3] and Florack [12] consider infinetely times differentiable convolution kernels which are rapidly decreasing functions in x, i.e. they are vanishing at ∞ faster than any inverse of polynomials.
  - Lindeberg uses kernels k_t which are Borel measurable in t [44], or kernels which converge for $t \to 0^+$ in the L^1 norm to the Dirac distribution [45].
  - Alvarez et al. [1] require that

    $$||T_t(f + hg) - (T_t(f) + hg)||_\infty \leq Cht$$

    for all $h, t \in [0, 1]$, and for all smooth f, g, where C may depend on f and g.
  - Pauwels et al. [52] assume that the convolution kernel $k_t(x)$ is separately continuous in x and in t.

- **Infinetesimal generator:**
  The existence of

  $$\lim_{t \to 0^+} \frac{T_t f - f}{t} =: A[f]$$

  guarantees that the semigroup can be represented by the evolution equation

  $$\partial_t u = A[u].$$

  From the mathematical literature it is well-known that the existence of an infinitesimal generator follows from the semigroup property when being combined with regularity assumptions [18].

- **Principle of maximum loss of figure impression:**
  See Section 4.1.

- **Causality:**
  The scale-space evolution should not create new level curves when increasing the scale parameter. If this is satisfied, iso-intensity linking through the scales is possible and a structure at a coarse scale can (in principle) be traced back to the original image.

  For this reason, Koenderink [41] required that at spatial extrema (with nonvanishing determinant of the Hessian) isophotes in scale-space are upwards convex; In 2-D he showed that at these extrema the diffusion equation

  $$\partial_t u = \alpha(x, t)\, \Delta u \tag{44}$$

  has to be satisfied. Hereby, α denotes a positive-valued function.

  Hummel [19] established the equivalence between causality and a maximum principle for certain parabolic operators.

  We may also derive the causality equation (44) and its N-dimensional generalizations by requiring that local extrema with positive or negative definite Hessians are not enhanced [3, 45]: This assumption states that such an extremum in x_0 at scale t_0 satisfies

  $$\begin{aligned} \partial_t u > 0 \qquad & \text{if } x_0 \text{ is a minimum,} \\ \partial_t u < 0 \qquad & \text{if } x_0 \text{ is a maximum.} \end{aligned}$$

This is just the causality requirement $\text{sign}(\partial_t u) = \text{sign}(\Delta u)$. Moreover, in 1-D, nonenhancement of local extrema is equivalent to the requirement that the number of local extrema does not increase [3, 44]. In higher dimensions, however, diffusion scale-spaces may create new extrema, see e.g. [68, 43, 8].

- **Nonnegativity:**
  If the nonnegativity of the convolution kernel,

$$k_t(x) \geq 0 \qquad \forall\, x, \quad \forall\, t > 0,$$

  is violated, new level crossings may appear for $t > 0$, such that the causality property does not hold.

  Within a linear framework with spatially continuous convolution kernels, nonnegativity is equivalent to the monotony requirement [1]

$$f(x) \;\leq\; g(x) \quad \forall\, x \quad \Longrightarrow \quad (T_t f)(x) \;\leq\; (T_t g)(x) \qquad \forall\, x, \quad \forall\, t > 0$$

  and the preservation of nonnegativity:

$$f(x) \;\geq\; 0 \quad \forall\, x \quad \Longrightarrow \quad (T_t f)(x) \;\geq\; 0 \qquad \forall\, x, \quad \forall\, t > 0.$$

- **Tikhonov regularization:**
  In the 1-D case, u is called a Tikhonov regularization of $f \in L^2(\mathbb{R})$, if it minimizes the energy functional

$$E_f[u] = \int\limits_{\mathbb{R}} \left[(f-u)^2 + \sum_{i=1}^{\infty} \lambda_i \left(\frac{d^i u}{dx^i} \right)^2 \right] dx \qquad (\lambda_i > 0).$$

  This concept and an N-dimensional generalization has been used by Nielsen, Florack and Deriche [48]. The first term under the integral ensures that u remains close to f, while the second one is responsible for the smoothness of u.

- **Average grey level invariance:**
  The average grey level invariance

$$\int\limits_{\mathbb{R}^N} T_t f \, dx = \int\limits_{\mathbb{R}^N} f \, dx \qquad \forall\, t > 0$$

  can be achieved by means of the continuity equation (28) in connection with reflecting or periodic boundary conditions. It boils down to the normalization condition

$$\int\limits_{\mathbb{R}^N} k_t(x) \, dx = 1,$$

  if we consider linear convolution kernels.

  In this case normalization is also equivalent to grey level shift invariance [1]:

$$\begin{aligned} T_t(0) &= 0, \\ T_t(f+C) &= T_t(f) + C \end{aligned}$$

  for all images f and for all constants C.

- **Flat kernel for $t \to \infty$:**
  For $t \to \infty$, one expects that the kernel spreads the information uniformly over the image. Therefore, if the integral over the kernel should remain finite, it follows that the kernel has to become entirely flat: $\lim\limits_{t \to \infty} k_t(x) = 0$.

- **Isometry invariance:**
Let $R \in \mathbb{R}^N$ be an orthogonal transformation (i.e. $\det R = \pm 1$) and define $(Rf)(x) := f(Rx)$. Then,
$$T_t(Rf) = R(T_t f) \qquad \forall\, f, \quad \forall\, t > 0.$$
In the 1-D case with a linear convolution kernel this invariance under rotation and mirroring comes down to the symmetry condition $k_t(x) = k_t(-x)$.

- **Homogeneity and isotropy:**
Koenderink [41] required that the scale-space treats all spatial points equally. He assumed that the diffusion equation (44), which results at extrema from the causality requirement, should be the same at each spatial position (regardless whether there is an extremum or not) and for all scales. He named these requirements homogeneity and isotropy[5].

- **Separability:**
The convolution kernel $k_t(x)$ with $x = (x_1, ..., x_N)^T \in \mathbb{R}^N$ may be split into N factors, each acting along one coordinate axis:
$$k_t(x) = k_{1,t}(x_1) \cdots k_{N,t}(x_N).$$

- **Scale invariance:**
Let $(S_\lambda f)(x) := f(\lambda x)$. Then there exists some $t'(\lambda, t)$ with
$$S_\lambda T_{t'} = T_t S_\lambda.$$
One may achieve this by requiring that, in the N-dimensional case, the convolution kernel k_t has the structure
$$k_t(x) = \frac{1}{\Psi^N(t)}\, \Phi\left(\frac{x}{\Psi(t)}\right)$$
with a continouus, strictly increasing rescaling function Ψ. This means that all kernels can be derived by stretching a parent kernel such that its area remains constant [52]. It is evident that this is related to the normalization condition.

  Scale invariance follows also from the semigroup property when being combined with isometry invariance and causality [45]. Moreover, scale invariance, translation invariace and isometry invariance result from the more general assumption of invariance under the spacetime symmetry group; see [12] for more details.

We observe that – despite the fact that all presented axiomatics use many similar requirements – not two of them are identical. Each of the 14 axiomatics confirms and enhances the evidence that the others give: that Gaussian scale-space is unique within a linear framework. This theoretical foundation is the backbone of a lot of successful applications of linear scale-space theory.
Nevertheless, apart from their historical merits, the early Japanese approaches differ from the well-known axiomatics after 1984 in several aspects:
Firstly, it is interesting to note that all Japanese axiomatics require only quite a few axioms in order to derive Gaussian scale-space. Even recent approaches which intend to use a minimal set of first principles do not utilize less axioms.
Iijima's 1-D and 2-D frameworks from 1959 and 1962, respectively, do not only belong to the most systematic derivations of Gaussian scale-space, they appear also rather modern: principles such as the semigroup property are typical for axiomatizations after 1990, and also the importance of scale invariance has been emphasized mainly in recent years [13, 52, 12].
Iijima's physical motivation for affine Gaussian scale-space from 1971 uses only two principles which reduce the essential features of linear diffusion filtering to a minimum. In this sense it is written in a similar spirit as Koenderink's derivation [41], which also uses two highly condensed requirements (although of very different nature). Concepts such as the principle of maximum loss of figure impression may remind some of the readers of properties of nonlinear scale-spaces like the Euclidean and affine shortening flow [1, 50, 56]: they shrink the Euclidean or affine perimeter of a closed curve as fast as

[5] In our terminology, homogeneity and isotropy are much stricter requirements than translation and isometry invariance. They enable Koenderink to derive Gaussian scale-space under only one additional assumption (causality).

possible. Moreover, the group-theoretical studies in [25] prove that Iijima has also pioneered modern scale-space analysis based on algebraic invariance theory such as in [1, 50, 54, 55].
Otsu's two-dimensional axiomatic is very appealing due to its simplicity: in contrast to many other approaches it does not require advanced mathematical techniques like Fourier analysis, complex integrals, or functional analysis in order to derive the uniqueness of the Gaussian kernel. It is therefore a well-suited approach even for undergraduate courses in image processing.

7 Discussion

In this paper we have analysed four axiomatics for the linear diffusion scale-space that have been unknown in the American and European image processing world. They reveal many interesting qualities which should trigger everyone who is interested in scale-space theory to have a closer look at them.
The discussed results demonstrate that an entire world of linear scale-space theory has evolved in Japan ranging from axiomatics for isotropic and affine Gaussian scale-space over algebraic invariance theory and deep structure analysis to hardware implementations for OCR. The Japanese scale-space paradigm was well-embedded into a general framework for pattern recognition and object classification [24, 33, 51], and many results have been established earlier than in America and Europe.
It is surprising that eastern and western scale-space theory have evolved with basically no interaction: To the best of our knowledge, the first citation of Iijima's work by non-Japanese scale-space researchers was made in 1996 [65]. Conversely, also Japanese work after 1983 was not always aware of American and European scale-space results. Some (English) papers by Makoto Sato's group [57, 58, 62, 61] cite both Iijima and Witkin. His paper with the probably most explicit reference to Iijima's work has been presented at the ICASSP '87 [57], where Sato and Wada cite the original Japanese versions of [25, 26] and state: *"The notion same as scale-space filtering was also proposed by T. Iijima in the field of pattern recognition. He derived a partial differential equation, called basic equation, from the continuity of the light energy in the waveform observation"*.
In 1992 several direct hints to Iijima's scale-space research can be found in the widespread journal *Proceedings of the IEEE* [47]. In an invited historical review of OCR methods written by Mori, Suen and Yanamoto, 10 out of the 193 key references were papers by Iijima. Concerning his contributions, the authors state that *"the concept of blurring was first introduced into pattern recognition by his work, whereas it was widely attributed to Marr in the West. Iijima's idea was derived from his study on modelling the vision observation system. (...) Setting reasonable conditions for the observation system, he proved that the mathematical form must be a convolution of a signal $f(x')$ with a Gaussian kernel"*. They also give a recommendation to the non-Japanese audience: *"Iijima's theory is not so easy to understand, but his recent book [34] is readable, although it is written in Japanese"*.
One can only speculate why nobody paid attention to these passages. Maybe, because none of the authors referred to one of Iijima's *English* scale-space papers. The present paper contains 10 references to publications by Iijima which are either originally written in English or available as complete English translations. They can be found in many libraries in America and Europe, and a short look at papers such as [24, 25] should convince everybody that there remains no justification to deny Iijima's pioneering role in linear scale-space theory because of language reasons.
Another reason might be that Iijima's work came too early to be appreciated: His theory was mathematically much more demanding than other methods at this time. At a stage where pattern recognition was still in its infancy and experimenting with very simple methods, it was not easy to make techniques popular, which are based on advanced mathematics. Also computing facilities were more restricted in the sixties and seventies than they are today. Despite very remarkable developments such as the scale-space based optical character readers, it was certainly more difficult to attract people by presenting computed results that demonstrate the advantages of a conceptually clean scale-space technique over ad-hoc strategies.
When scale-space became popular in America and Europe in the eighties, the situation was different: Computing power was much higher, and the pattern recognition and computer vision community had experienced sufficiently many frustrations with ad hoc methods to get aware of their limitations and to become mature for better founded techniques which take advantage of centuries of research in mathematics and physics. Today it is possible to establish an international conference solely devoted to scale-space ideas which attracts people from many countries and disciplines [17]. Would this have been possible 30 years ago? Certainly not.

Unfortunately, it seems that Iijima was not the only one who has pioneered the field of partial differential equations in image processing far ahead of his time, so that his work fell into oblivion for decades. Another example is the fact that already in 1965 the Nobel prize winner Dennis Gabor – the inventor of optical holography and the so-called Gabor functions – proposed a deblurring algorithm based on combining mean curvature flow with backward smoothing along flowlines [15, 46]. This long-time forgotten method is similar to modern PDE techniques for image enhancement.
Maybe this review helps a little bit that the pioneering work of these people receives the acknowledgement that it deserves. It would also be nice if it encourages the mutual interest between the Japanese and the western scale-space community. The fact that the same theory has been developed twice in two very different cultures shows that this theory is natural and that it is worthwhile to study all of its various aspects.

Acknowledgements. We thank Alfons Salden, Luc Florack, Peter Johansen and Jon Sporring for helpful comments and stimulating discussions. Parts of this work have been carried out while S.I. and J.W. were with the Image Sciences Institute of Utrecht University, The Netherlands. It has also been supported by the Danish National Science Research Council under grant no. 9502164.

References

[1] L. Alvarez, F. Guichard, P.-L. Lions, J.-M. Morel, *Axioms and fundamental equations in image processing*, Arch. Rational Mech. Anal., Vol. 123, 199–257, 1993.

[2] Y. Aoki, T. Iijima, *Pluralizing method of simple similarity*, IEICE Trans. Inf. Syst., Vol. E79-D, 485–490, 1996 (in English).

[3] J. Babaud, A.P. Witkin, M. Baudin, R.O. Duda, *Uniqueness of the Gaussian kernel for scale space filtering*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 8, 26–33, 1986.

[4] R. van den Boomgaard, *The morphological equivalent of the Gauss convolution*, Nieuw Archief Voor Wiskunde, Vierde Serie, Deel 10, 219–236, 1992.

[5] R.W. Brockett, P. Maragos, *Evolution equations for continuous-scale morphological filtering*, IEEE Trans. Signal Processing, Vol. 42, 3377–3386, 1994.

[6] A.M. Bruckstein, D. Shaked, *On projective invariant smoothing and evolution of planar curves and polygons*, CIS Report No. 9328, Computer Science Dept., Technion, Haifa 32000, Israel, 1993.

[7] S. Chinveeraphan, R. Takamatsu, M. Sato, *A hierarchical description of digital grayscale images based on image dipoles*, Proc. 13th Int. Conf. Pattern Recognition (ICPR 13, Vienna, Aug. 25–30, 1996), Vol. B, 246–250, 1996.

[8] J. Damon, *Local Morse theory for solutions to the heat equation and Gaussian blurring*, J. Differential Equations, Vol. 115, 368–401, 1995.

[9] R. Deriche, O. Faugeras, *Les EDP en traitement des images et vision par ordinateur*, Traitement du Signal, Vol. 13., No. 6, 1997.

[10] F. Dibos, *Analyse multiéchelle invariante par transformations projectives*, C. R. Acad. Sci. Paris, t. 320, Série I, 917–921, 1995.

[11] O. Faugeras, *Sur l'évolution de courbes simples du plan projectif réel*, C. R. Acad. Sci. Paris, t. 317, Série I, 565–570, 1993.

[12] L.M.J. Florack, *Data, models, and images*, Proc. 1996 IEEE Int. Conf. Image Processing (ICIP–96, Lausanne, Sept. 16–19, 1996), Vol. I, 469–472, 1996.

[13] L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, M.A. Viergever, *Scale and the differential structure of images*, Image Vision Comput., Vol. 10, 376–388, 1992.

[14] J. Fourier, *Théorie analytique de la chaleur*, Didot, Paris, 1822.

[15] D. Gabor, *Information theory in electron microscopy*, Laboratory Investigation, Vol. 14, 801–807, 1965.

[16] B.M. ter Haar Romeny (Ed.), *Geometry-driven diffusion in computer vision*, Kluwer, Dordrecht, 1994.

[17] B. ter Haar Romeny, L. Florack, J. Koenderink, M. Viergever (Eds.), Scale-Space Theory in Computer Vision, Lecture Notes in Comp. Science, Vol. 1252, Springer, Berlin, 1997.

[18] E. Hille, R.S. Philips, *Functional analysis and semi-groups*, American Mathematical Society, Providence, 1957.

[19] R.A. Hummel, *Representations based on zero-crossings in scale space*, Proc. IEEE Comp. Soc. Conf. Computer Vision and Pattern Recognition (CVPR '86, Miami Beach, June 22–26, 1986), IEEE Computer Society Press, Washington, 204–209, 1986.

[20] T. Iijima, *Basic theory of pattern observation*, Papers of Technical Group on Automata and Automatic Control, IECE, Japan, Dec. 1959 (in Japanese).

[21] T. Iijima, *Basic theory on normalization of pattern (In case of typical one-dimensional pattern)*, Bulletin of the Electrotechnical Laboratory, Vol. 26, 368–388, 1962 (in Japanese).

[22] T. Iijima, *Observation theory of two-dimensional visual patterns*, Papers of Technical Group on Automata and Automatic Control, IECE, Japan, Oct. 1962 (in Japanese).

[23] T. Iijima, *Basic theory on normalization of two-dimensional visual pattern*, Studies on Information and Control, Pattern Recognition Issue, IECE, Japan, No. 1, 15–22, 1963 (in Japanese).

[24] T. Iijima, *Theory of pattern recognition*, Electronics and Communications in Japan, 123–134, Nov. 1963 (in English).

[25] T. Iijima, *Basic equation of figure and observational transformation*, Systems, Computers, Controls, Vol. 2, No. 4, 70–77, 1971 (in English).

[26] T. Iijima, *Basic theory on the construction of figure space*, Systems, Computers, Controls, Vol. 2, No. 5, 51–57, 1971 (in English).

[27] T. Iijima, *A suppression kernel of a figure and its mathematical characteristics*, Systems, Computers, Controls, Vol. 2, No. 6, 16–23, 1971 (in English).

[28] T. Iijima, *Basic theory on normalization of figure*, Electronics and Communications in Japan, Vol 54–C, No. 12, 106–112, 1971 (in English).

[29] T. Iijima, *A system of fundamental functions in an abstract figure space*, Systems, Computers, Controls, Vol. 1.2, No. 6, 96–103, 1971 (in English).

[30] T. Iijima, *Basic theory on feature extraction of figures*, Systems, Computers, Controls, Vol. 3, No. 1, 32–39, 1971 (in English).

[31] T. Iijima, *Basic theory on the structural recognition of a figure*, Transactions of the Institute of Electronics and Communication Engineers of Japan, English Abstracts, Vol. 55-D, No. 8, 28–29, 1972.

[32] T. Iijima, *Theoretical studies on the figure identification by pattern matching*, Transactions of the Institute of Electronics and Communication Engineers of Japan, English Abstracts, Vol. 55-D, No. 8, 29–30, 1972.

[33] T. Iijima, *Pattern recognition*, Corona-sha, 1973 (in Japanese).

[34] T. Iijima, *Theory of pattern recognition*, Series of Basic Information Technology, Vol. 6, Morishita Publishing, 1989 (in Japanese).

[35] T. Iijima, H. Genchi, K. Mori, *A theoretical study of pattern identification by matching method*, Proc. First USA–Japan Computer Conference (Tokyo, Oct. 3–5, 1972), 42–48, 1972 (in English).

[36] T. Iijima, H. Genchi, K. Mori, *A theory of character recognition by pattern matching method*, Proc. First Int. Joint Conf. Pattern Recognition (IJCPR, Washington, Oct. 1973), 50–56, 1973 (in English).

[37] A. Imiya, R. Katsuta, *Extraction of a structure feature from three-dimensional objects by scale-space analysis*, B. ter Haar Romeny, L. Florack, J. Koenderink, M. Viergever (Eds.), Scale-space theory in computer vision, Lecture Notes in Comp. Science, Vol. 1252, Springer, Berlin, 345–348, 1997.

[38] P.T. Jackway, M. Deriche, *Scale-space properties of the multiscale morphological dilation–erosion*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 18, 38–51, 1996.

[39] G. Kin, M. Sato, *Scale space filtering on spherical pattern*, Proc. 11th Int. Conf. Pattern Recognition (ICPR 11, The Hague, Aug. 30 – Sept. 3, 1992), Vol. C, 638–641, 1992.

[40] B.B. Kimia, A. Tannenbaum, S.W. Zucker, *Toward a computational theory of shape: An overview*, O. Faugeras (Ed.), Computer vision – ECCV '90, Lecture Notes in Comp. Science, Vol. 427, Springer, Berlin, 402–407, 1990.

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

[42] J.J. Koenderink, *Scale in perspective*, J. Sporring, M. Nielsen, L. Florack, P. Johansen (Eds.), Gaussian scale-space theory, Kluwer, Dordrecht, xv–xx, 1997.

[43] L.M. Lifshitz, S.M. Pizer, *A multiresolution hierarchical approach to image segmentation based on intensity extrema*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 12, 529–540, 1990.

[44] T. Lindeberg, *Scale-space for discrete signals*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 12, 234–254, 1990.

[45] T. Lindeberg, *On the axiomatic formulations of linear scale-space*, J. Sporring, M. Nielsen, L. Florack, P. Johansen (Eds.), Gaussian scale-space theory, Kluwer, Dordrecht, 75–97, 1997.

[46] M. Lindenbaum, M. Fischer, A. Bruckstein, *On Gabor's contribution to image enhancement*, Pattern Recognition, Vol. 27, 1–8, 1994.

[47] S. Mori, C.Y. Suen, K. Yanamoto, *Historical review of OCR research and development*, Proc. IEEE, Vol. 80, 1029–1058, 1992.

[48] M. Nielsen, L. Florack, R. Deriche, *Regularization, scale-space, and edge detection filters*, B. Baxton, R. Cipolla (Eds.), Computer vision – ECCV '96: Volume II, Lecture Notes in Comp. Science, Vol. 1065, Springer, Berlin, 70–81, 1996.

[49] H. Nishino, *PIPS (pattern information processing system) project – background and outline*, Proc. 4th Int. Joint Conf. Pattern Recognition (IJCPR 4, Kyoto, Nov. 7–10, 1978), 1152–1161, 1978.

[50] P.J. Olver, G. Sapiro, A. Tannenbaum, *Classification and uniqueness of invariant geometric flows*, C. R. Acad. Sci. Paris, t. 319, Série I, 339–344, 1994.

[51] N. Otsu, *Mathematical studies on feature extraction in pattern recognition*, Ph.D. thesis, Report No. 818, Electrotechnical Laboratory, 1-1-4, Umezono, Sakura-mura, Niihari-gun, Ibaraki, Japan, 1981 (in Japanese).

[52] E.J. Pauwels, L.J. Van Gool, P. Fiddelaers, T. Moons, *An extended class of scale-invariant and recursive scale space filters*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 17, 691–701, 1995.

[53] P. Perona, J. Malik, *Scale space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 12, 629–639, 1990.

[54] N. Rougon, F. Prêteux, *Understanding the structure of diffusive scale-spaces*, M.-O. Berger, R. Deriche, I. Herlin, J. Jaffré, J.-M. Morel (Eds.), ICAOS '96: Images, wavelets and PDEs, Lecture Notes in Control and Information Sciences, Vol. 219, Springer, London, 103–110, 1996.

[55] A.H. Salden, *Dynamic scale-space paradigms*, Ph.D. thesis, Utrecht University, The Netherlands, 1996.

[56] G. Sapiro, A. Tannenbaum, *Affine invariant scale-space*, Int. J. Comput. Vision, Vol. 11, 25–44, 1993.

[57] M. Sato, T. Wada, *A hierarchical representation of random waveforms by scale-space filtering*, Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP–87, Dallas, April 1987), 273–276, 1987.

[58] M. Sato, T. Wada, H. Kawarada, *A morphological study on structure line*, Proc. 9th Int. Conf. Pattern Recognition (ICPR 9, Rome, Nov. 14–17, 1988), 559–562, 1988.

[59] J. Sporring, M. Nielsen, L. Florack, P. Johansen (Eds.), Gaussian scale-space theory, Kluwer, Dordrecht, 1997.

[60] J.L. Stansfield, *Conclusions from the commodity expert project*, A.I. Memo 671, Artificial Intelligence Lab., Massachusetts Inst. of Technolgy, Cambridge, MA, U.S.A., 1980.

[61] T. Wada, Y.H. Gu, M. Sato, *Scale-space filtering for periodic waveforms*, Systems and Computers in Japan, Vol. 22, 45–54, 1991.

[62] T. Wada, M. Sato, *Scale-space tree and its hierarchy*, Proc. 10th Int. Conf. Pattern Recognition (ICPR 10, Atlantic City, June 16–21, 1990), IEEE Computer Society Press, Los Alamitos, Vol. 3, 103–108, 1990.

[63] J. Weickert, *Nonlinear diffusion scale-spaces: From the continuous to the discrete setting*, M.-O. Berger, R. Deriche, I. Herlin, J. Jaffré, J.-M. Morel (Eds.), ICAOS '96: Images, wavelets and PDEs, Lecture Notes in Control and Information Sciences, Vol. 219, Springer, London, 111–118, 1996.

[64] J. Weickert, *Anisotropic diffusion in image processing*, Teubner Verlag, Stuttgart, 1997, to appear.

[65] J.A. Weickert, B.M. ter Haar Romeny, M.A. Viergever, *Conservative image transformations with restoration and scale-space properties*, Proc. 1996 IEEE Int. Conf. Image Processing (ICIP–96, Lausanne, Sept. 16–19, 1996), Vol. I, 465–468, 1996.

[66] J. Weickert, S. Ishikawa, A. Imiya, *On the history of Gaussian scale-space axiomatics*, J. Sporring, M. Nielsen, L. Florack, P. Johansen (Eds.), Gaussian scale-space theory, Kluwer, Dordrecht, 45–59, 1997.

[67] A.P. Witkin, *Scale-space filtering*, Proc. Eighth Int. Joint Conf. on Artificial Intelligence (IJCAI '83, Karlsruhe, Aug. 8–12, 1983), Vol. 2, 1019–1022, 1983.

[68] A.L. Yuille, T.A. Poggio, *Scaling theorems for zero crossings*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 8, 15–25, 1986.

[69] N.-Y. Zhao, *Feature extraction of images by stable gaze tree*, Ph.D. thesis, Dept. of Computer Science, Tokyo Inst. of Technology, 1985 (in Japanese).

[70] N.-Y. Zhao, T. Iijima, *Theory on the method of determination of view-point and field of vision during observation and measurement of figure*, IECE Japan, Trans. D, Vol. J68-D, 508–514, 1985 (in Japanese).

[71] N.-Y. Zhao, T. Iijima, *A theory of feature extraction by the tree of stable view-points*, IECE Japan, Trans. D, Vol. J68-D, 1125–1132, 1985 (in Japanese).