International Journal of Approximate Reasoning 51 (2010) 1129–1144

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Possibilistic signal processing: How to handle noise?

Kevin Loquin [a,*], Olivier Strauss [b], Jean-Francois Crouzet [c]

[a] IRIT, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse Cedex 9, France
[b] LIRMM, Université Montpellier II, 161, rue Ada, F-34392 Montpellier Cedex 5, France
[c] I3M, Université Montpellier II, Place Eugene Bataillon, F-34095 Montpellier, France

ARTICLE INFO

Article history:
Available online 24 August 2010

Keywords:
Signal processing
Kernel methods
Possibility distribution
Noise quantization
Choquet integral

ABSTRACT

We propose a novel approach for noise quantifier at each location of a signal. This method is based on replacing the conventional kernel-based approach extensively used in signal processing by an approach involving another kind of kernel: a possibility distribution. Such an approach leads to interval-valued resulting methods instead of point-valued ones. We propose a theoretical justification to this approach and we show, on real and artificial data sets, that the length of the obtained interval and the local noise level are highly correlated. This method is non-parametric and has an advantage over other methods since no assumption about the nature of the noise has to be made, except its local ergodicity. Besides, the propagation of the noise in the involved signal processing method is direct and does not require any additional computation.

1. Introduction

Reliability of most signal processing methods inherently depends on the possibility of adjusting their parameters to account for noise level over the input signal. Examples of such procedures are edge detection [16], motion estimation [1], segmentation [19], shape-from-shading [33], where noise and the feature to be extracted (edge, motion, object) can be confused. Noise level estimation is also very important for (obviously) denoising [25,26] or anti-aliasing methods, as well as for signal enhancement methods like super-resolution [11], sensor fusion [2,28] algorithms or image restoration.

Noise in a signal is usually referred to random variations of its measurements. These variations can be produced by several factors including thermal effect, saturation, sampling, quantization and transmission. Since repeating the acquisition process is usually not possible, the noise level has to be estimated by means of a single signal occurrence.

Random noise level is generally considered as being independent from the signal level and modelled by a random quantity added to the signal. One of the most widely encountered models assumes this random noise as being centered and normally distributed. However, phenomena like film grain, speckle, impulse noise, sampling effect, quantization or saturation induce a fluctuation of the signal that cannot be modelled by a Gaussian zero mean process. For example, in medical images produced by a gamma camera, the noise is rather described by a Poisson process (i.e. the noise level depends on the signal level). Indeed, with a gamma camera, the acquisition consists of counting the gamma photons detected by each sensor. The greater the accumulated quantity of gamma photons, the greater the potential noise level.

In early approaches (see e.g. [24]), noise estimation consists in assuming stationarity of the random variations of the signal. The computation of the standard deviation of the noise is performed by analyzing the signal obtained by high-pass filtering of the original signal. The main challenge in these estimations is to be able to tell whether high-frequency signal variations are due to noise or to the signal itself.

* Corresponding author.
E-mail addresses: loquin@irit.fr (K. Loquin), strauss@lirmm.fr (O. Strauss), Jean-Francois.Crouzet@univ-montp2.fr (J.-F. Crouzet).

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2010.08.004

In more recent papers, some authors propose to abandon either stationarity or additivity of the noise. Rangayyan et al. [27] consider an adaptive neighbourhood approach that is able to account for an additive non-stationary noise. Corner et al. show that analyzing the Laplacian of the signal allows to deal with both additive and multiplicative noise [4].

Unfortunately, neither additive nor multiplicative random noise are good models for real signal contamination, even for instance, for conventional CCD (Charge-Coupled Device) sensor [16]. Therefore, many approaches [13,16,20] propose to model the acquisition noise as being Poisson distributed.

In these model-based approaches, noise is assumed to follow a hypothetically known distribution and noise level estimation consists in estimating the different parameters on which the variance of the assumed distribution depends. Moreover, any model-based method assumes the acquisition system to be error calibrated.

If nothing can be assumed about the nature of the noise, except its local ergodicity, only a very local approach has to be considered to estimate the noise level for each location or, at least, for each user-selected homogeneous region of the signal. Moreover, since signal processing mainly consists of extracting or estimating some physically meaningful characteristics from intensity values of the signal, it should be important to understand how the uncertainty due to random perturbation propagates through any algorithm step.

A wide range of those signal processing methods relies on a kernel-based approach [17] for direct or iterative, linear or non-linear algorithms and for filtering (stochastic, band pass, anti-aliasing, etc.), geometrical transformations (rescaling, rotations, homographies, anamorphosis, etc.), sampling rate conversion, fusion, for enhancing or removing details, etc. The kernels usually encountered are probability distributions: a kernel is a positive function whose total weight (its integral in the infinite domain and its sum in the finite domain) equals 1. The main difficulty in these kernel-based methods is that the nature of both signal and perturbation can change during the complete analysis, from step to step.

By switching from probability theory to possibility theory, we propose new methods that account for a lack of knowledge about the proper kernel to be used [18]. The lack of knowledge is handled by the fact that a possibility distribution represents a convex hull of probability distributions and hence of kernels. In the proposed adaptation of the usual kernel methods, the conventional Lebesgue integral operator is replaced by a pair of Choquet integrals according to the possibility measure and the necessity measure associated with the chosen possibility distribution. The resulting interval (and more precisely its length) reflects the lack of knowledge of the modeller on the most adequate kernel to use.

As an example, the use of the interval-valued gradient estimation of an image, proposed in [15], leads to a threshold-free robust edge detector. This robustness is due to the fact that the length of the interval-valued estimation is highly correlated with the input image random noise. The information (about the noise) contained in the resulting interval is properly taken into account in the edge detector, thus enabling an automatic rejection of the "false" edges due to noise.

In this paper, we propose to study the link between the length of the interval-valued output of a possibilistic filter and the random noise level of the input signal. Actually, we discuss the fact that the spirit of this approach is, to our opinion, better founded than the usual noise level estimators. Furthermore, we expose some research tracks to theoretically justify our approach and we propose to highlight the empirical correlation between the length of the interval-valued output of a possibilistic filter and the random noise of the input signal on real and simulated repetitions of image acquisitions.

The paper is organized as follows. In Section 2, we present an extension of kernel-based signal processing methods based on possibility distributions. We present, among other methods, the possibilistic filtering approach that allows the computation of a convex family of filtering operations. We theoretically justify this extension by means of Theorem 1. In Section 3, we describe our method for locally estimating the noise level at each location of a signal. We propose, in this section, a theoretical study of this estimator. In Section 4, we compare our method to three other usual noise level estimates on synthetic and real noisy images, before concluding in Section 5.

2. A possibilistic extension of kernel-based signal processing

2.1. Kernel-based signal processing

The kernel-based signal processing framework can be equally applied to discrete or continuous signals. Whatever the nature of the signal, we work on an underlying continuous domain $\Omega = \mathbb{R}$. A discrete signal is defined for indices of $k \in \mathbb{Z}$ representing locations ω_k of Ω, i.e. a discrete signal S is given by $S = (S_k)_{k\in\mathbb{Z}}$. A continuous signal is defined for all the elements of Ω, i.e. a continuous signal S is given by $S = \{S(u) : u \in \Omega\}$.

In case of a continuous signal, the kernel-based signal processing operation on a location $\omega \in \Omega$ is defined by the convolution operation:

$$\widehat{S}(\omega) = \int_\Omega S(u)\kappa(\omega - u)du.$$

For any $\omega \in \Omega$, $\kappa^\omega = \{\kappa(\omega - u) : u \in \Omega\}$ is the continuous convolution kernel κ shifted to the location ω of Ω. $\widehat{S}(\omega)$ is thus obtained by:

$$\widehat{S}(\omega) = \int_\Omega S(u)\kappa^\omega(u)du. \tag{1}$$

The kernel-based signal processing operation (1) transforms a continuous signal S into a continuous signal $\widehat{S}$. When this operation is performed for a discrete set of locations identified by their indices $n \in \mathbb{Z}$, it can be expressed by:

$$\widehat{S}_n = \int_\Omega S(u)\kappa^n(u)du. \tag{2}$$

For any $n \in \mathbb{Z}, \kappa^n$ is a continuous kernel. The kernel-based signal processing operation (2) transforms a continuous signal S into a discrete signal $\widehat{S}$ by means of a set of continuous kernels $(\kappa^n)_{n\in\mathbb{Z}}$.

In case of a discrete input signal, the discrete kernel-based signal processing operation is defined, for a location $\omega \in \Omega$, by:

$$\widehat{S}(\omega) = \sum_{k\in\mathbb{Z}} S_k \kappa_k^\omega. \tag{3}$$

For any $\omega \in \Omega$, $\kappa^\omega = (\kappa_k^\omega)_{k\in\mathbb{Z}}$ is a discrete kernel. The kernel-based signal processing operation (3) transforms a discrete signal $(S_k)_{k\in\mathbb{Z}}$ into a continuous signal $\widehat{S}$ by means of a set of discrete kernels $\{\kappa^\omega : \omega \in \Omega\}$. When the location $\omega \in \Omega$ matches one of the sampling locations of S, i.e. when there is $n \in \mathbb{Z}$, such that $\omega = \omega_n$, then the discrete kernel-based signal processing operation can be written as a discrete convolution operation:

$$\widehat{S}_n = \sum_{k\in\mathbb{Z}} S_k \kappa_k^n, \tag{4}$$

where $\kappa^n = (\kappa_k^n)_{k\in\mathbb{Z}} = (\kappa_{n-k})_{i\in\mathbb{Z}}$ is the discrete convolution kernel κ shifted to the location ω_n of Ω. The kernel-based signal processing operation (4) transforms a discrete signal $(S_k)_{k\in\mathbb{Z}}$ into a discrete signal $(\widehat{S}_n)_{n\in\mathbb{Z}}$ by means of a set of discrete kernels $(\kappa^n)_{n\in\mathbb{Z}}$.

In many applications like low-pass filtering, signal sampling or signal interpolation, the used convolution kernels are positive and have a unitary sum or integral, i.e.

$$\sum_{i\in\mathbb{Z}} \kappa_i = 1, \quad \text{for a discrete kernel,}$$
$$\int_\Omega \kappa(u)du = 1, \quad \text{for a continuous kernel.}$$

We call this condition “summativity”. A convolution kernel which fulfils this condition is called a summative kernel.

A summative convolution kernel can be seen as a probability distribution that induces a probability measure P_κ, computed by:

$$\forall A \subseteq \mathbb{Z}, \quad P_\kappa(A) = \sum_{i\in A} \kappa_i, \quad \text{for a discrete kernel,}$$
$$\forall A \subseteq \Omega, \quad P_\kappa(A) = \int_A \kappa(u)du, \quad \text{for a continuous kernel.}$$

For any summative kernel κ, the shifted convolution kernels κ^n or κ^ω are still summative kernels. Thus, in case of a summative convolution kernel, expressions (1)–(4) can be reformulated as expectations of the signal S considering the probability measure P_{κ^n} or P_{κ^ω}, i.e.

$$\widehat{S}_n = \mathbb{E}_{P_{\kappa^n}}(S), \quad \text{for obtaining a discrete signal,} \tag{5}$$
$$\widehat{S}(\omega) = \mathbb{E}_{P_{\kappa^\omega}}(S), \quad \text{for obtaining a continuous signal.} \tag{6}$$

Note that P_{κ^n} or P_{κ^ω} can both be equally discrete or continuous probability measures. The used expectation operator is a discrete sum if the probability measure is discrete and is an integral if the probability measure is continuous.

A naive interpretation of expressions (5) and (6) is that the processed signal could be seen as the expected value of the signal, knowing that the uncertainty concerning the location is modelled by the probability measure P_{κ^n} or P_{κ^ω}. However, in most encountered cases, this naive interpretation is not relevant. For instance, when the processing consists of filtering a signal, generally, its aim is not to evaluate the most likely value of the real signal under uncertainty modelled by the probability P_{κ^n} or P_{κ^ω}, but to modify (for instance derivate or remove the high-frequency part of) the input signal according to the practitioner’s needs. The only reason why we propose to reformulate kernel-based processing with the expectation operator is that it enables us to switch from the usual probability theory to imprecise probability theory and thus to work with a family of summative convolution kernels.

2.2. Possibilistic extension of kernel-based signal processing

By expressing the kernel-based signal processing operators as expectation operators (5) and (6) according to probability measures, we open new perspectives to this approach. We can explore adaptations of this operator by means of new uncertainty theories. This development has similarities with the development of the theory of aggregation operators [22]. Instead of using an additive measure, i.e. a probability measure, to represent the neighbourhood of a particular location, we propose to use the simple non-additive confidence measure called a possibility measure [8]. We propose to use this theory among others because of its computational simplicity. First, a possibility measure is a tool that can be simply modelled by a function, its associated distribution (which is a set of weights in the discrete case) on Ω, whereas most of the other imprecise probability theories [32] require more assessments. Besides, we propose to use the Choquet integral, that extends the usual expectation operator to possibility measures. Using the Choquet integral leads to low-computational algorithms.

This section presents, interprets and theoretically justifies this new approach, based on possibility measures and Choquet integrals. The core idea of this approach is to process a signal by means of a family of convolution kernels.

2.2.1. A possibility distribution is a family of summative kernels

A possibility measure Π is a non-additive confidence measure [6]. It possesses a dual confidence measure, called a necessity measure, denoted by N and computed in this way:

$$\forall A \subseteq \Omega \text{ (or } \mathbb{Z} \text{ in the discrete case)}, \quad N(A) = 1 - \Pi(A^c). \tag{7}$$

The two measures, Π and N, encode a family of probability measures, denoted by $\mathcal{M}(\Pi)$, and defined by:

$$\mathcal{M}(\Pi) = \{P | \forall A \subseteq \Omega \text{ (or } \mathbb{Z}),\ N(A) \leqslant P(A) \leqslant \Pi(A)\}.$$

This encoding property is due to the sensitivity analysis interpretation [31] of possibility theory.

As mentioned before, a possibility measure can be defined from a possibility distribution π. Such a distribution is normalized in the sense that:

$$\max_{i\in\mathbb{Z}} \pi_i = 1, \quad \text{in the discrete case,}$$
$$\max_{\omega\in\Omega} \pi(\omega) = 1, \quad \text{in the continuous case.}$$

This condition can also be called maxitivity, by analogy to the term summativity. Its associated possibility measure is obtained by:

$$\forall A \subseteq \mathbb{Z}, \quad \Pi_\pi(A) = \max_{i\in A} \pi_i, \quad \text{in the discrete case,}$$
$$\forall A \subseteq \Omega, \quad \Pi_\pi(A) = \max_{\omega\in A} \pi(\omega), \quad \text{in the continuous case.}$$

Thus a unique possibility distribution π can encode a whole family of convolution kernels κ with unitary gain, denoted by $\mathcal{M}(\pi)$ and defined by:

$$\mathcal{M}(\pi) = \{\kappa | \forall A \subseteq \Omega \text{ (or } \mathbb{Z}),\ N_\pi(A) \leqslant P_\kappa(A) \leqslant \Pi_\pi(A)\}.$$

As shown in [17], this family of convolution kernels is capable of representing a lack of knowledge about the resolution of the proper kernel to be used in different applications. The resolution of a kernel, seen as a neighbourhood, characterizes its concentration on a point. In fact, one of the main properties of the family $\mathcal{M}(\pi)$ is as follows: if a kernel κ_Δ, with a resolution Δ, belongs to $\mathcal{M}(\pi)$ then any kernel κ_δ, obtained by dilation from the same basic kernel κ, with a resolution $\delta \leqslant \Delta$, belongs to $\mathcal{M}(\pi)$. This property can be very useful for modelling sampling or interpolation processes since, for these kinds of applications, only the maximal resolution of the convolution kernel can usually be set.

This family of convolution kernels being defined, the extension of the convolution (or expectation) operator follows naturally.

2.2.2. The possibilistic extension of kernel-based signal processing

We propose to denote by $\mathbb{C}_v(f)$ the Choquet integral of a bounded function f according to a capacity v which is a non-additive measure [6]. The Choquet integral can be considered as a generalization of the conventional expectation operator when the involved measure is additive. In other words, if $v = P$ is a probability measure, then $\mathbb{C}_v = \mathbb{E}_P$.

Since a possibility measure is a capacity, the conventional expectation operator has to be replaced by the Choquet integral [5]. Using a Choquet integral and a possibility distribution leads to an interval-valued expectation, instead of a single-valued one. In the discrete case, the upper and lower bounds of this interval-valued expectation are respectively given by:

$$\begin{cases} \overline{S}_n = \mathbb{C}_{\Pi_{\pi^n}}(S), \\ \underline{S}_n = \mathbb{C}_{N_{\pi^n}}(S). \end{cases} \tag{8}$$

In the continuous case, the upper and lower bounds of this interval-valued expectation are respectively given by:

$$\begin{cases} \overline{S}(\omega) = \mathbb{C}_{\Pi_{\pi^\omega}}(S), \\ \underline{S}(\omega) = \mathbb{C}_{N_{\pi^\omega}}(S). \end{cases} \tag{9}$$

The key point of this approach is that the interval-valued expectation obtained by means of a possibility distribution is the set of all the single-valued expectations obtained by using all the convolution kernels encoded by the considered possibility distribution. This property holds for both the discrete and continuous case. However, the ways to prove it are different.

As a preliminary to the theorem (and its proof) justifying this property in the discrete case, some notations are necessary. Let us denote by $\mathcal{L}(\mathbb{Z})$ the set of bounded sets of weights on $\mathbb{Z}$, i.e. $\forall u = (u_i)_{i\in\mathbb{Z}} \in \mathcal{L}(\mathbb{Z}),\ \max_{i\in\mathbb{Z}}|u_i| < \infty$. In [31], this set is called the set of bounded gambles on $\mathbb{Z}$. Denote $\mathcal{B}(\mathbb{Z})$, the set of binary (i.e. $\{0,1\}$-valued) sets of weights on $\mathbb{Z}$. Obviously, $\mathcal{B}(\mathbb{Z}) \subset \mathcal{L}(\mathbb{Z})$. $\mathcal{B}(\mathbb{Z})$ can be seen as the set of events on $\mathbb{Z}$.

Theorem 1. *Let π^ω be a discrete possibility distribution. $\forall S \in \mathcal{L}(\mathbb{Z}),\ \forall \kappa^\omega \in \mathcal{M}(\pi^\omega)$,*

$$\mathbb{C}_{N_{\pi^\omega}}(S) \leqslant \mathbb{E}_{P_{\kappa^\omega}}(S) \leqslant \mathbb{C}_{\Pi_{\pi^\omega}}(S). \tag{10}$$

Moreover, the bounds are reached: $\forall S \in \mathcal{L}(\mathbb{Z})$, $\exists \kappa_1^\omega, \kappa_2^\omega \in \mathcal{M}(\pi^\omega)$, such that,

$$\mathbb{C}_{N_{\pi^\omega}}(S) = \mathbb{E}_{P_{\kappa_1^\omega}}(S),$$
$$\mathbb{C}_{\Pi_{\pi^\omega}}(S) = \mathbb{E}_{P_{\kappa_2^\omega}}(S).$$

Proof. The natural extension principle [31] is required to prove Theorem 1. Note that the natural extension of a probability measure P, defined for all the events A of $\mathcal{B}(\mathbb{Z})$, is the expectation according to P, defined for all S of $\mathcal{L}(\mathbb{Z})$. Similarly, the natural extension of a possibility measure Π, defined for all the events A of $\mathcal{B}(\mathbb{Z})$, is the Choquet integral with respect to Π, defined for all S of $\mathcal{L}(\mathbb{Z})$. This remark is true for the more general belief functions.

The natural extension, as defined by Walley, is conservative concerning the imprecision of a possibility measure. The family of natural extensions of the probability measures of the family $\mathcal{M}(\pi^\omega)$, denoted by $E(\mathcal{M}(\pi^\omega))$, is the same as the family of expectations dominated by the Choquet integral according to π^ω, noted $\mathcal{M}(\mathbb{C}_{\Pi_{\pi^\omega}})$. This property of the natural extension can be found in Walley's book [31] for an upper prevision $\overline{P}$ and its associated set of linear previsions $\mathcal{M}(\overline{P})$. It is enough to conclude that $\forall S \in \mathcal{L}(\mathbb{Z}),\ \forall \kappa^\omega \in \mathcal{M}(\pi^\omega)$,

$$\mathbb{C}_{N_{\pi^\omega}}(S) = \min_{\kappa^\omega \in \mathcal{M}(\pi^\omega)} \mathbb{E}_{P_{\kappa^\omega}}(S), \tag{11}$$
$$\mathbb{C}_{\Pi_{\pi^\omega}}(S) = \max_{\kappa^\omega \in \mathcal{M}(\pi^\omega)} \mathbb{E}_{P_{\kappa^\omega}}(S). \qquad \square \tag{12}$$

This theorem is also valid for continuous signals.

Theorem 2. *Let π^ω be a continuous possibility distribution. For any bounded continuous signal S, $\forall \kappa^\omega \in \mathcal{M}(\pi^\omega)$,*

$$\mathbb{C}_{N_{\pi^\omega}}(S) \leqslant \mathbb{E}_{P_{\kappa^\omega}}(S) \leqslant \mathbb{C}_{\Pi_{\pi^\omega}}(S). \tag{13}$$

Moreover, the bounds are reached. For any bounded continuous signal S, $\exists \kappa_1^\omega, \kappa_2^\omega \in \mathcal{M}(\pi^\omega)$, such that,

$$\mathbb{C}_{N_{\pi^\omega}}(S) = \mathbb{E}_{P_{\kappa_1^\omega}}(S),$$
$$\mathbb{C}_{\Pi_{\pi^\omega}}(S) = \mathbb{E}_{P_{\kappa_2^\omega}}(S).$$

The proof is derived from domination theorems proved by Denneberg [6, Proposition 10.3] and Schmeidler [29, Proposition 3].

Note that in the case of a discrete processing of a positive discrete signal S (which is the case of the images that will be processed in Section 4), the Choquet integrals, forming the upper and lower expectations, can be explicitly computed by:

$$\overline{S}_n = \mathbb{C}_{\Pi_{\pi^n}}(S) = \sum_{i\in\mathbb{Z}} \Pi_{\pi^n}(A_{(i)})(S_{(i)} - S_{(i-1)}), \tag{14}$$
$$\underline{S}_n = \mathbb{C}_{N_{\pi^n}}(S) = \sum_{i\in\mathbb{Z}} N_{\pi^n}(A_{(i)})(S_{(i)} - S_{(i-1)}). \tag{15}$$

The index notation $(\cdot)$ indicates a permutation that sorts the sample locations such that $S_{(1)} \leqslant S_{(2)} \leqslant \cdots \leqslant S_{(N)}$ and $A_{(i)}$ is the set of samples locations whose value is greater than $S_{(i)}$, i.e. $A_{(i)} = \{j \in \mathbb{Z} | S_j > S_{(i)}\}$. By convention, $S_{(0)} = 0$.

2.2.3. How to choose the possibility distribution?

The use of a possibility distribution as a family of convolution kernels is new in signal processing. This approach does not offer clues (especially to possibility theory novices) for choosing the possibility distribution that matches the practitioner's knowledge on the proper convolution kernel to be used. The choice procedures that we propose in this section depend on the nature of the information held by the practitioner. In all these procedures, the nature of the meta-information about the lack of knowledge about the proper convolution kernel to be used depends on the context.

Objective probability/possibility transformation. Let us consider the situation where the practitioner knows the expression of a most plausible proper kernel function and its maximal resolution. This situation is very common in any continuous to discrete scheme. Let κ be this kernel. In that case, we propose to base the assessment of the possibility distribution on the objective probability/possibility transformation of Dubois et al. [7,9,10]. It is defined by:

$$\forall i \in \mathbb{Z}, \quad \pi_i^o = 1 - P_\kappa(I_i), \quad \text{in the discrete case,}$$
$$\forall \omega \in \Omega, \quad \pi^o(\omega) = 1 - P_\kappa(I_\omega), \quad \text{in the continuous case,}$$

where

$$\forall i \in \mathbb{Z}, \quad I_i = \{j \in \mathbb{Z} : \kappa_j \geqslant \kappa_i\}, \quad \text{in the discrete case,}$$
$$\forall \omega \in \Omega, \quad I_\omega = \{x \in \Omega : \kappa(x) \geqslant \kappa(\omega)\}, \quad \text{in the continuous case.}$$

It can be proved that this objective transformation based on the confidence intervals satisfies three basic principles:

1. Possibility–probability coherence: one should select a possibility distribution coherent with κ, i.e. such that $\kappa \in \mathcal{M}(\pi^o)$.
2. Ordinal faithfulness: the chosen possibility distribution should preserve the ordering of elementary events, namely, $\pi^o(u) > \pi^o(v)$ if and only if $\kappa(u) > \kappa(v)$ (the same condition holds for the discrete case).
3. Informativity: the information content of π^o is maximized so as to preserve as much from κ as possible. It means that π^o is the most specific possibility distribution dominating κ.

This last point is very significant to us. This objective transformation should be advocated when the practitioner is highly confident in his knowledge of κ since the obtained possibility distribution is the most specific distribution dominating κ.

Besides, this most specific possibility distribution obtained from κ dominates all the summative kernels whose resolution is smaller than the resolution of κ. This property has been proved in [17] for the granularity measure which is an index of resolution of summative kernels.

Subjective probability/possibility transformation. Let us consider the situation where the practitioner knows an expression of a most plausible proper kernel but has a vague idea of the upper resolution of his kernel. In that case, the practitioner's confidence on his choice is weaker than in the objective transformation case. We propose to use the subjective transformation [7,9,10]. This transformation results in a less specific possibility distribution than the objective transformation. It is defined by:

$$\forall i \in \mathbb{Z}, \quad \pi_i^s = \sum_{j \in \mathbb{Z}} \min(\kappa_j, \kappa_i), \quad \text{in the discrete case,}$$
$$\forall \omega \in \Omega, \quad \pi^s(\omega) = \int_\Omega \min(\kappa(x), \kappa(\omega))dx, \quad \text{in the continuous case.}$$

It can easily be shown [17] that

$$\forall i \in \mathbb{Z}, \quad \pi_i^s = \pi_i^o + \#(I_i)\kappa_i, \quad \text{in the discrete case,}$$
$$\forall \omega \in \Omega, \quad \pi^s(\omega) = \pi^o(\omega) + \mu(I_\omega)\kappa(\omega), \quad \text{in the continuous case,}$$

where #() is the cardinality and $\mu()$ is the Lebesgue measure. Those expressions show that $\pi^s \geqslant \pi^o$, i.e. that π^s is less specific than π^o.

Summarizing, probability/possibility transformations studied by Dubois et al. [7,9,10] can be used when the practitioner has a vague idea of the convolution kernel to be used. The objective transformation results in the smallest family containing the original kernel. The subjective transformation results in a less specific family than the objective transformation. The choice of a transformation depends on the confidence that the practitioner has on his judgement and more precisely on what he knows about the resolution of the kernel to be used. When he knows the upper bound of its resolution, we advocate the use of the objective transformation. When he doubts this upper bound, we advocate the subjective transformation.

Chebyshev's inequality. Let us consider, in the continuous case only, the situation where the practitioner only knows the two first moments of the kernel to be used. In that case, we advocate the use of a probabilistic inequality to construct a family of probability distributions which forms a possibility distribution.

Dubois et al. [7,10] proposed the possibility distribution, denoted $\pi_{(m,\sigma)}$, defined from the Chebyshev's inequality by:

$$\pi_{(m,\sigma)}(m - \omega\sigma) = \pi_{(m,\sigma)}(m + \omega\sigma) = \min\left(1, \frac{1}{\omega^2}\right), \quad \forall \omega > 0.$$

This possibility distribution is symmetric and encodes, among others, all the probability distributions of mean m and of variance σ^2.

For information, procedures for assessing a symmetric possibility distribution from very few measurements have been proposed by Mauris [23].

Triangular possibility distribution. Let us consider the situation where the practitioner only knows that the summative convolution kernel is symmetric and has a given bounded support. In that case, we propose to use the triangular possibility distribution defined by:

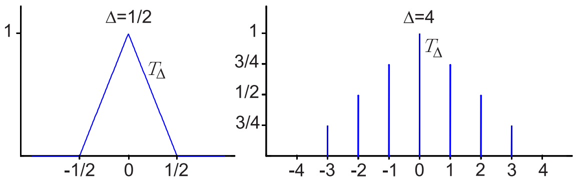

Fig. 1. Triangular possibility distributions for $\Delta = \frac{1}{2}$ (continuous) and $\Delta = 4$ (discrete).

$$\text{for } \Delta \in \mathbb{N},\ \forall i \in \mathbb{Z},\ T_{\Delta i} = \max\left(0, 1 - \frac{|i|}{\Delta}\right),\ \text{in the discrete case,} \tag{16}$$

$$\text{for } \Delta \in \mathbb{R}^+,\ \forall u \in \Omega,\ T_\Delta(u) = \max\left(0, 1 - \frac{|u|}{\Delta}\right),\ \text{in the continuous case.} \tag{17}$$

The notation $|\cdot|$ corresponds to the absolute value.

We propose to use a triangular possibility distribution since, as shown in [10], for a spread Δ, its associated triangular distribution encodes (among others) all the symmetric convolution kernels with a support whose length is smaller (or equal) than twice its spread, i.e. 2Δ. Fig. 1 shows illustrative examples of a continuous (on the left) and a discrete (on the right) triangular possibility distribution.

3. Noise level estimation via possibilistic filtering

3.1. Some existing noise level estimators

In signal processing, noise can be seen as a disturbance that affects the signal. This disturbance is usually due to external factors that distort the signal acquisition, transmission or storage. More often than not, noise is considered as a statistical variation of the signal value. Within this interpretation, noise can be associated to a probability distribution reflecting an impossible experiment consisting in performing many (ideally an infinity) acquisitions in exactly the same conditions. In practice, noise has to be quantified from few (and even only one) experiments.

Usually, noise quantization is based on a real or assumed pre-knowledge about its probabilistic model. The noise model (i.e. the statistical fluctuations of the signal value) at a particular location is described by its two first moments, i.e. its mean and standard deviation, or, more generally, an aggregated value (mean, mode, median, etc.) and a spread factor (interquartile, standard deviation, etc.). In this view, it is important to distinguish between noise level estimation and estimation of the spread factor parameter of the noise probabilistic model. Indeed, the spread factor estimation is a particular case of noise level estimation, which consists of quantifying the noise with a relative index, called the noise level, independent of any chosen model or any unit of noise measurement. By relative index, we mean that the value taken by a single occurrence of a noise level estimate is generally meaningless. However, such estimators can be used as tools for comparing the noise level between different signals or between different locations of the same signal.

3.1.1. Nugget effect

In geostatistics, a very interesting way for catching the noise level is employed. Geostatistic is a branch of applied statistics that concentrates on the description of spatial patterns [3,12,21]. The central tool of geostatistic is the random function which describes the uncertainty of a given spatial characteristic over a domain. The structural assumption underlying most of the geostatistical methods is based on the intuitive idea that, the closer are the regions of interest, the more similar are their associated characteristic values.

However, this intuitive idea is no more so obvious when looking at the closest pairs of sample locations of a spatial data set. Indeed, in general, when plotting the empirical increments of a particular observed property, function of the distance between different sample locations (i.e. plotting the sample variogram [3]), these increments do not vanish when the distance tends to 0 (see Fig. 2). This discontinuity, which is supposedly due to geostatistical noise, is called the "nugget effect". This denomination comes from the fact that in gold deposits, gold commonly occurs as nuggets of pure gold that are much smaller than the size of a sample.

When translating this concept from geostatistics to signal processing, the nuggets effect can be illustrated as follows: the variability of a subset A of the signal domain is supposed to reflect the co-occurrence of the intrinsic local variability of the supposed continuous signal underlying the samples and a measurement error. This measurement error sums up the systematic error due to the impulse response of the sensor, the imprecision due to sampling and quantization of the signal and a random variability due to noise. Typically, the variability due to the signal increases with the radius Δ of the subset A. On the contrary, the variability due to the noise is usually supposed not to depend on Δ. This assumption is reasonable when the sampling is regular and the random noise is supposed to be locally stationary. Thus, if $A_\Delta(\omega)$ is a neighbourhood of radius Δ of the location ω, $V(A_\Delta(\omega))$, a measure of the variability of $A_\Delta(\omega)$ is such that:

$$\lim_{\Delta \to 0} V(A_\Delta(\omega)) = v(\omega) \tag{18}$$

with $\nu(\omega)$ being the noise level due to measurement error at location ω. This limit relates to the nuggets effect in the geostatistic field [12]. However, due to sampling, $\nu(\omega)$ cannot be computed because the local variability cannot be estimated for a scale smaller than the sampling distance h.

A standard technique for catching this noise level is to plot a variogram, i.e. to plot the variability of all the sampling locations, $V(A_\Delta)$, as a function of Δ, the separation between the sampling locations. A manual fitting is generally required to provide an estimation of the noise level, which is the value of the regression curve for the radius $\Delta = 0$. This estimation is denoted by v. Sometimes, an automatic fitting procedure, as least squares regression, is performed, but this is not recommended by the geostatisticians since it does not permit to take into account their additional pieces of information about the studied field.

However, this method presupposes that the noise is stationary all over the signal. Moreover, the choice to be made for a particular variogram model is not generally justified in signal processing. The expert's knowledge is generally not available in this scientific domain to evaluate local dependencies, whereas in geostatistic, the expert, according to the physical nature of the studied area, can provide such information.

3.1.2. Local noise level estimators

A more point-wise estimation of this noise level can be obtained by means of a small neighbourhood around each location of the underlying continuous domain Ω of the studied signal. This approach is based on assuming local ergodicity. Local ergodicity states that the local variability of the signal in a small neighbourhood of a location reflects the statistical variations of the signal at this location, due to noise. When the studied signal is discrete, such a neighbourhood can usually be represented by a discrete probability distribution defined over the set of sampling locations by $\kappa^\omega = (\kappa_i^\omega)_{i\in\mathbb{Z}}$. When the studied signal is continuous, such a neighbourhood can usually be represented by a continuous probability distribution defined on Ω by $\kappa^\omega = \{\kappa^\omega(u): u \in \Omega\}$. The computation of the noise level leads to a weighted sum or integral due to the additivity of the probability measure. Estimation of the noise level is given by:

$$v^2(\omega) = \sqrt{\sum_{i\in\mathbb{Z}} (S_i - \widehat{S}(\omega))^2 \kappa_i^\omega}, \tag{19}$$

if variability is measured by the standard deviation. And by:

$$v^1(\omega) = \sum_{i\in\mathbb{Z}} |S_i - \widehat{S}(\omega)| \kappa_i^\omega, \tag{20}$$

if variability is measured by the mean error. $\widehat{S}(\omega)$ is the expectation of the signal S at location ω, i.e.

$$\widehat{S}(\omega) = \sum_{i\in\mathbb{Z}} S_k \kappa_i^\omega. \tag{21}$$

If the signal to be analyzed is continuous, the noise affecting the signal S at location ω can be estimated by means of the continuous counterparts of the expressions (19)–(21), given by:

$$\begin{aligned}
v^2(\omega) &= \sqrt{\int_\Omega (S(u) - \widehat{S}(\omega))^2 \kappa^\omega(u) du}, \\
v^1(\omega) &= \int_\Omega |S(u) - \widehat{S}(\omega)| \kappa^\omega(u) du, \\
\widehat{S}(\omega) &= \int_\Omega S(u) \kappa^\omega(u) du
\end{aligned}$$

for a continuous summative neighbourhood κ^ω.

Most of the kernels used to perform these estimations are uni-modal, centered and symmetric around the location ω.

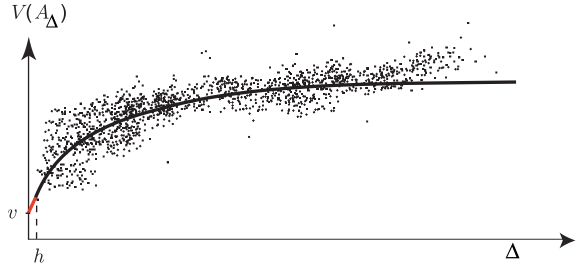

Fig. 2. Qualitative example of variogram.

3.2. Principle of our possibilistic noise level estimator

Some classes of linear filters, like low-pass filters, can be directly modelled by a kernel-based signal processing operation. In that case, the impulse response of the filter is the convolution kernel.

Let $S = (S_i)_{i\in\mathbb{Z}}$ be a discrete signal. Processing S by a filter, defined by its impulse response κ, mathematically corresponds to the discrete convolution of S by κ. The value $\widehat{S}_n$ of the filtered signal at the nth location is thus obtained by:

$$\widehat{S}_n = \sum_{i\in\mathbb{Z}} S_i \kappa_{n-i}.$$

This discrete linear filtering operator and the following one are exactly the same as the discrete kernel-based signal processing operator (4):

$$\widehat{S}_n = \sum_{i\in\mathbb{Z}} S_i \kappa_i^n.$$

In the continuous case, processing a continuous signal $S = \{S(u):u \in \Omega\}$ defined on Ω by a continuous filter, defined by its impulse response κ, mathematically corresponds to the convolution of S by κ. The value $\widehat{S}(\omega)$ of the filtered signal at the location ω of Ω is thus obtained by:

$$\widehat{S}(\omega) = \int_\Omega S(u)\kappa(\omega - u)du.$$

This continuous linear filtering operator and the following one are exactly the same as the continuous kernel-based signal processing operator (1):

$$\widehat{S}(\omega) = \int_\Omega S(u)\kappa^\omega(u)du.$$

Generally, low-pass filters have impulse responses (convolution kernels) that are positive and have a unitary gain in order to avoid attenuation or amplification of the input signal. Most low-pass filters are smoothing or averaging filters. The impulse responses of such filters are summative kernels.

What we aim at illustrating in the following is the relevance of the noise quantification ability of the possibility-based filtering.

Suppose one low-pass filters a signal with two different filters having the same cut-off frequency f_c. Such filters eliminate from the input signal any component having frequency higher than the cut-off frequency f_c (this explains the denomination "low-pass filter"). Suppose that the maximal frequency of the input signal is lower than f_c. Then the two output signals will be approximately equal. Now, suppose that one applies these same filtering procedures to an input signal with components having frequencies beyond f_c. Then, generally, the output signals will be different. The more different are the shapes of both convolution kernels, the more likely different are the outputs of the two filters.

Now, consider the same procedure with a family of low-pass filters (instead of just two). The previous remark still holds. Moreover, the dispersion in the outputs of this family of low-pass filters is a direct consequence of the high-frequency level of the input signal. If we now suppose that the high frequencies of the input signal are only due to noise, then the dispersion in the outputs of this family of low-pass filters can be considered as a marker of the variability of the input signal due to noise.

As mentioned before, the impulse responses of the usual linear low-pass filters are summative convolution kernels (uniform, Gaussian filters, etc.). Since a possibility distribution represents a convex family of summative convolution kernels, we propose to replace the usual low-pass filtering based on a convolution kernel by a possibility distribution-based low-pass filtering procedure.

The imprecision of the outputs of a possibility distribution-based filter is quantified by the length of the interval $[\underline{S}_n, \overline{S}_n]$, in a discrete filtering context (resp. $[\underline{S}(\omega), \overline{S}(\omega)]$ in a continuous filtering context), as defined by expressions (8) (resp. (9)).

In the following, we will theoretically justify and empirically illustrate the conjecture that this imprecision can be seen as a marker of the noise level of the input signal at the considered location.

Actually, under the assumption of local ergodicity, we propose to estimate the noise level by:

$$\lambda_n = \overline{S}_n - \underline{S}_n \quad \text{in the discrete case, at location } n, \tag{22}$$

$$\lambda(\omega) = \overline{S}(\omega) - \underline{S}(\omega) \quad \text{in the continuous case, at location } \omega. \tag{23}$$

As the most usual low-pass filters have impulse responses which are uni-modal and symmetric convolution kernels around n (or ω), the triangular possibility distribution plays a central role in possibility-distribution-based filtering. Indeed, as already mentioned, the triangular possibility distribution is the most specific possibility distribution that dominates the class of all uni-modal symmetric convolution kernels with the same mode and support.

In discrete signal processing, in order to weaken the influence of the signal variations on our noise level estimator, we have to choose a highly specific possibilistic neighbourhood. We define this neighbourhood as being the most specific one that dominates any probabilistic neighbourhood that can be used as an interpolator. Since the family of these interpolation kernels is the family of centered kernels whose support is included in $[-1,1]$, the triangular neighbourhood with this support is the best candidate. Its discrete expression can be simply represented by the vector:

$$T_2 = \begin{pmatrix} 0.5 \\ 1 \\ 0.5 \end{pmatrix}. \tag{24}$$

In the case of image processing, i.e. with a 2D discrete signal, the used triangular (or pyramidal) neighbourhood of each pixel can be simply represented by the possibilistic 3×3 matrix:

$$T_{2\times 2} = \begin{pmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{pmatrix}. \tag{25}$$

3.3. Theoretical justifications: some tracks

The aim of this section is to propose some preliminary theoretical justifications for the possibilistic noise level estimator. In this scope, we can notice that the only way to theoretically prove that a noise level estimator is relevant is to compare, on the one hand a measure of the statistical variations of an infinite number of signal acquisitions, and on the other hand, the proposed estimator. However, this approach is infeasible because of the infinity. If we restrict this procedure to a finite number of signal acquisitions (which is feasible), the obtained comparison becomes empirical. Indeed, infinity is the link between the theoretical concept of objective probability (and of associated measures of variations) and the reality that probability aims at catching. Regarding these remarks, we come up with the conclusion that, if no assumption is made on the noise model, it is not possible to theoretically study a noise level estimator.

Our justification proposal involves continuous signals and a white Gaussian additive noise. Let S, a measured signal, be the sum of an underlying f_c-smooth signal s and of an *approximated Gaussian white noise η of level σ*, i.e.

$$\forall \omega \in \Omega, \quad S(\omega) = s(\omega) + \eta(\omega). \tag{26}$$

Let us define here what we mean under the terms f_c-smooth signal and *approximated Gaussian white noise η of level σ*.

Definition 1. An f_c-smooth signal s has a Fourier transform, denoted by $F[s]$, whose support is included in the set $[-f_c f_c]$.

Definition 2. An approximated Gaussian white noise η of level σ is such that

- $\forall \omega \in \Omega$, $\eta(\omega)$ is a centered Gaussian random variable (i.e. with a zero mean), and
- its autocorrelation function is such that $\eta \otimes \eta(\omega) = \int_\Omega \eta(u)\eta(u-\omega)du = \sigma^2\delta_0(\omega)$, where δ_0 is a finite energy approximation of the Dirac δ distribution.

The set of signals S that are a combination of an underlying *f_c-smooth signal s* and an *approximated Gaussian white noise η of level σ* is denoted by $\mathcal{S}(f_c, \sigma)$.

The theoretical justification that we propose do not work directly with the whole family $\mathcal{M}(\pi)$, but with a subset of $\mathcal{M}(\pi)$. For defining this sub-family, let us consider another family of filters, denoted by $\mathcal{F}(f_c)$, that we call *f_c-smoothing filters*.

Definition 3. An f_c-smoothing filter κ has a Fourier transform, denoted by $F[\kappa]$, which is equal to *1*, at least for all the frequencies under f_c, i.e. $F[\kappa](\xi) = 1$, $\forall |\xi| \leqslant f_c$.

First, it should be noted that the definition of an f_c-smoothing filter is consistent with the concept of low-pass filter. Indeed, the convolution operation of any signal g by the impulse response of a filter κ, which is the filtering procedure, given by $(g * \kappa)(\omega)$, $\forall \omega \in \Omega$, can be replaced by a product operation in the space of the Fourier transforms, i.e. $F[g](\xi)F[\kappa](\xi)$, for all ξ in the frequency domain. A low-pass filter is supposed not to alter the signal for low frequencies, i.e. for frequencies $|\xi| \leqslant f_c$. Moreover, if $F[\kappa](\xi) = 1$, for all $|\xi| \leqslant f_c$, then $F[g](\xi)F[\kappa](\xi) = F[g](\xi)$. In other words, the low frequency part of g is not altered as is expected with low-pass filters.

This restriction allows us to work with a family of summative kernels whose convolution with an f_c-smooth signal is involutive:

Proposition 1. *For any f_c-smooth signal s and for all κ of $\mathcal{F}(f_c)$, $s * \kappa = s$.*

Proof. Indeed, for frequencies $|\xi| \leqslant f_c$, $F[s](\xi)F[\kappa](\xi) = F[s](\xi)$, because $F[\kappa](\xi) = 1$ and for frequencies $|\xi| > f_c$, $F[s](\xi)F[\kappa](\xi) = 0 = F[s](\xi)$, because $F[s](\xi) = 0$. □

For the sake of completeness, the following proposition has to be proven:

Proposition 2. *$\mathcal{F}(f_c)$ is a family of filters whose impulse responses are summative kernels.*

Proof. Let κ be the impulse response of a filter of $\mathcal{F}(f_c)$. Its Fourier transform can be written as $F[\kappa](\xi) = \int_\Omega \kappa(u)e^{-2\pi j u\xi}du$. Therefore, $F[\kappa](0) = \int_\Omega \kappa(u)du$. Since $F[\kappa](0) = 1$, it follows that any kernel of $\mathcal{F}(f_c)$ is summative. □

Our trick is to work with the family $\mathcal{M}(\pi) \cap \mathcal{F}(f_c) \subseteq \mathcal{M}(\pi)$, i.e. with the family of filters of $\mathcal{M}(\pi)$ who are f_c-smoothing filters. Therefore, the obtained interval with this family of filters is necessarily included in the obtained interval with the family $\mathcal{M}(\pi)$, given by $[\underline{S}(\omega), \overline{S}(\omega)]$. Indeed,

$$\overline{S}(\omega) = \mathbb{C}_{\Pi_{\pi^\omega}}(S) = \max_{\kappa^\omega \in \mathcal{M}(\pi^\omega)} \mathbb{E}_{P_{\kappa^\omega}}(S) = \max_{\kappa \in \mathcal{M}(\pi)} (S * \kappa)(\omega) \geqslant \max_{\kappa \in \mathcal{M}(\pi) \cap \mathcal{F}(f_c)} (S * \kappa)(\omega),$$
$$\underline{S}(\omega) = \mathbb{C}_{N_{\pi^\omega}}(S) = \min_{\kappa^\omega \in \mathcal{M}(\pi^\omega)} \mathbb{E}_{P_{\kappa^\omega}}(S) = \min_{\kappa \in \mathcal{M}(\pi)} (S * \kappa)(\omega) \leqslant \min_{\kappa \in \mathcal{M}(\pi) \cap \mathcal{F}(f_c)} (S * \kappa)(\omega).$$

Thus, we have that $[\underline{S}'(\omega), \overline{S}'(\omega)] \subseteq [\underline{S}(\omega), \overline{S}(\omega)]$, where $\underline{S}'(\omega)$ and $\overline{S}'(\omega)$ are respectively the lower and upper bounds of the outputs of the family of filters $\mathcal{M}(\pi) \cap \mathcal{F}(f_c)$ at location ω. Formally,

$$\overline{S}'(\omega) = \max_{\kappa \in \mathcal{M}(\pi) \cap \mathcal{F}(f_c)} (S * \kappa)(\omega),$$
$$\underline{S}'(\omega) = \min_{\kappa \in \mathcal{M}(\pi) \cap \mathcal{F}(f_c)} (S * \kappa)(\omega).$$

Now, for a given maxitive kernel π, let $D(\omega)$ denote, $\forall \omega \in \Omega$, the difference between the upper and the lower bounds of the set of outputs that we obtain with the filters of the family of summative kernels $\mathcal{F}(f_c) \cap \mathcal{M}(\pi)$, i.e.

$$D(\omega) = \overline{S}'(\omega) - \underline{S}'(\omega). \tag{27}$$

It can be noted that $D(\omega)$ is always smaller than our noise level estimate $\lambda(\omega)$.

Considering all these preliminaries, let us expose the main point of this theoretical justification of our approach:

Theorem 3. *Let π be a maxitive kernel. Let $\sigma > 0$ and $f_c > 0$ be chosen. There exists a constant $\alpha = \alpha(\pi, f_c)$, such that for any signal S of $\mathcal{S}(f_c, \sigma)$, we have,*

$$\mathbb{E}(D(\omega)) = \alpha(\pi, f_c)\sigma, \quad \forall \omega \in \Omega. \tag{28}$$

Proof. First, from Proposition 1, we can deduce that $\forall \kappa \in \mathcal{F}(f_c)$ (and thus $\forall \kappa \in \mathcal{M}(\pi) \cap \mathcal{F}(f_c)$),

$$S * \kappa = (s + \eta) * \kappa = s * \kappa + \eta * \kappa = s + \eta * \kappa. \tag{29}$$

Thus, $D(\omega)$ simplifies as follows:

$$D(\omega) = \max_{\kappa \in \mathcal{M}(\pi) \cap \mathcal{F}(f_c)} (\eta * \kappa)(\omega) - \min_{\kappa \in \mathcal{M}(\pi) \cap \mathcal{F}(f_c)} (\eta * \kappa)(\omega),$$

because the component s is eliminated by subtraction.

We aim at proving that $D(\omega)$ has an expected value proportional to σ (28) with a coefficient of proportionality α which depends on π and f_c.

Let us now write $\eta = \sigma\eta^*$, where η^* is an approximated Gaussian white noise with unit standard deviation. By linearity we get $D = \sigma D^*$ with

$$D^*(\omega) = \max_{\kappa \in \mathcal{M}(\pi) \cap \mathcal{F}(f_c)} (\eta^* * \kappa)(\omega) - \min_{\kappa \in \mathcal{M}(\pi) \cap \mathcal{F}(f_c)} (\eta^* * \kappa)(\omega).$$

As η is a stationary process, η^* is also stationary, and the fact that D^* is a function of η^* preserves this stationarity.

Thus expression (28) holds, where $\alpha(\pi, f_c) = \mathbb{E}(D^*(\omega))$, the coefficient of proportionality depends on π and f_c, but not on ω due to the stationarity of D^*. □

By showing the proportionality of the expected value of the difference between the upper and lower outputs obtained with the family $\mathcal{M}(\pi) \cap \mathcal{F}(f_c)$, with the noise level σ, we can ensure that the intervals $[\underline{S}'(\omega), \overline{S}'(\omega)]$ increase with σ. The inclusion of $[\underline{S}'(\omega), \overline{S}'(\omega)]$ in the interval $[\underline{S}(\omega), \overline{S}(\omega)]$ used in our noise level estimator $\lambda(\omega)$ thus guarantees that the noise level σ will act on our estimator. Indeed, when σ increases, the length of $[\underline{S}'(\omega), \overline{S}'(\omega)]$ and thus the length of $[\underline{S}(\omega), \overline{S}(\omega)]$ increases and finally, $\lambda(\omega)$ increases.

The same theoretical results can be shown for discrete signals and filters.

4. Experiments

4.1. Experiment on an image affected by a simulated noise

For this first experiment, we synthesized a set of noisy images from the benchmark image Lena. A Gaussian noise is simulated for standard deviations ranging from 0 to 60 and added to the original Lena image (cf. Fig. 3). With this set of noisy

Fig. 3. Images of Lena [14] with simulated Gaussian noise with standard deviations of 0, 30 and 60.

images, we can directly compare the noise level estimates presented in this paper (19), (20) and (22) with the simulated added noise.

This experiment attempts to show the ability of the possibility distribution-based approach, presented in subsection 3.2, to quantify the noise level on an image when the noise is supposed to be locally ergodic. The noise level is known and represented by the standard deviation of the added approximated Gaussian white noise.

The average over all the pixels of the noisy images of the noise level estimates (19), (20) and (22) is plotted on Fig. 4 versus the level of the simulated added noise. The highest curve corresponds to the standard deviation estimate, i.e. expression (19) with a 3×3 convolution kernel, the curve in the middle, corresponds to the mean error estimate, i.e. expression (20) with a 3×3 convolution kernel and the lowest curve corresponds to the possibility distribution-based noise level estimate, i.e. expression (22).

As can be seen in Fig. 4, all these estimators are good markers of the noise level, since the three plotted curves seem to fit affine functions of the noise level. The part of the curves with small simulated noise levels (i.e. with standard deviation lower than 5) is not fully in agreement with this remark. This is due to the fact that for low noise levels, the signal to noise ratio is high and the observed variations of the noisy image are mainly due to the image, and not to the noise.

From this experiment, we cannot pretend that our estimator is better than the other existing local estimators to quantify the noise level, since the three curves are very similar. However, put in a more general context, our approach looks more appropriate to handle the noise in further processing. In any usual method, an additional step is necessary to handle the noise in the processing. The advantage of the possibilistic approach is that noise level quantization is part of the processing (in that case the filtering) of the data without any additional computation.

4.2. Experiment on real images with real noise

A Hoffman 2D Brain Phantom (Data Spectrum Corporation), denoted by HBP, was filled with a $99m$ technetium solution ($148MBq/L$) and placed in front of one of the detectors of a dual-head gamma camera using a low-energy high-resolution parallel-hole collimator (INFINIA, General Electric Healthcare). A dynamic study was performed to provide 1000 planar acquisitions (acquisition time: 1 s; average count per image 1.5 kcounts, 128×128 images to satisfy the Shannon condition), representing 1000 measures of a random 2D image supposedly ruled by a Poisson process.

The acquisition time being very short, the images are very noisy, i.e. the signal to noise ratio is very low. More precisely, the average pixel value in the brain corresponds to a coefficient of variation of the Poisson noise of 69%. $I_{n,p}$ is the measured activity of the nth pixel within the pth acquired image. Note that Fig. 5 only shows the 40×35 central parts of the images that contains the HBP projection.

This experiment attempts to show that the possibility distribution-based noise level estimator (22) is more correlated to the statistical variations of the image than the standard deviation noise estimation approach.

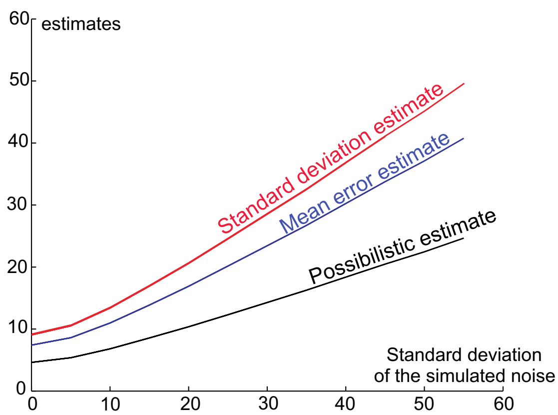

Fig. 4. Usual and possibilistic local estimates of the noise level.

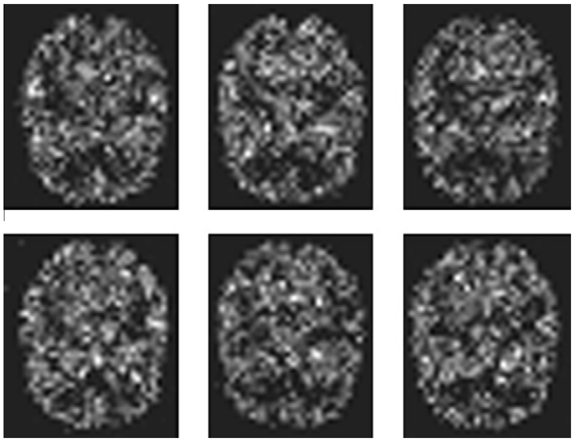

Fig. 5. Six images from the 1000 HBP direct acquisitions.

The randomness of the radioactive decay being statistically described by the Poisson probability, it cannot really be assumed to be stationary all over the image. Since the signal to noise ratio is very low, the local variation of the activity level, in the neighbourhood of each pixel, is still highly correlated with the statistical variations due to acquisition noise.

On the one hand, the statistical variation of the activity of the nth pixel can be estimated by its standard deviation σ_n all over its different realizations:

$$\sigma_n = \sqrt{\frac{1}{999}\sum_{p=1}^{1000}(I_{n,p} - m_n)^2}, \tag{30}$$

with m_n, the weighted mean of the image at the nth pixel:

$$m_n = \frac{1}{1000}\sum_{p=1}^{1000} I_{n,p}. \tag{31}$$

On the other hand, the local variation of the measurement in the neighbourhood of the nth pixel within the pth image can be estimated by computing the standard deviation via the expression (19) with a highly specific kernel (the same experiment made with expression (20) led to similar results). In this experiment, we propose two estimates of this standard deviation: $\gamma_{n,p}$ is computed by using a 3×3 uniform neighbourhood, and $\delta_{n,p}$ is computed by using a Gaussian kernel with a standard deviation equal to 1.6, i.e. a kernel whose bandwidth has been adapted to equal the bandwidth of the uniform kernel [17,30].

In the meantime, we compute, for each image, an interval-valued activity $[\underline{I}_{n,p}, \overline{I}_{n,p}]$ by using the possibility distribution-based method described in Section 3.2. The local variation in the neighbourhood of the nth pixel within the pth image is estimated by the length $\lambda_{n,p}$ of each interval:

$$\lambda_{n,p} = \overline{I}_{n,p} - \underline{I}_{n,p}. \tag{32}$$

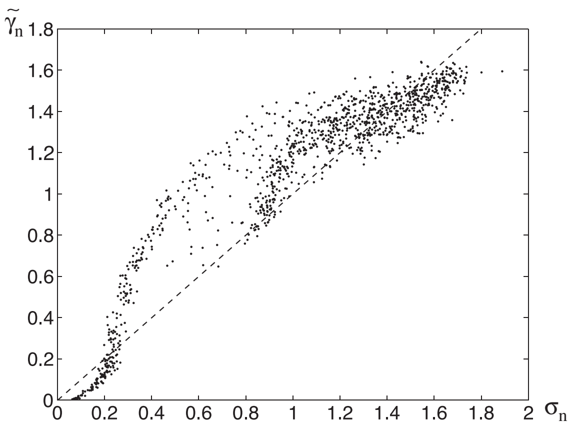

Fig. 6. Local variation measured by using a 3×3 uniform kernel versus the statistical variation.

We aim at testing whether the distribution of the estimated standard deviation σ_n is correlated or not with $\gamma_{n,p}$, $\delta_{n,p}$ and $\lambda_{n,p}$. To provide a clear illustration, we compute, for each n, the mean of the distributions of the deviation measures: $\tilde{\gamma}_n = \frac{1}{1000}\sum_{p=1}^{1000}\gamma_{n,p}$, $\tilde{\delta}_n = \frac{1}{1000}\sum_{p=1}^{1000}\delta_{n,p}$ and $\tilde{\lambda}_n = \frac{1}{1000}\sum_{p=1}^{1000}\lambda_{n,p}$.

Fig. 6 plots $\tilde{\gamma}_n$ versus σ_n, as well as the straight line $\sigma_n = \tilde{\gamma}_n$, Fig. 7 plots $\tilde{\delta}_n$ versus σ_n, as well as the straight line $\sigma_n = \tilde{\delta}_n$ and Fig. 8 plots $\tilde{\lambda}_n$ versus σ_n, as well as the straight line $\sigma_n = \tilde{\lambda}_n$.

These figures clearly show that all these estimations are, on average, correlated with σ_n. The choice of the value 1.6 for the Gaussian kernel is appropriate since the estimated local standard deviations $\tilde{\delta}_n$ are in the same range as the statistical standard deviations σ_n. Indeed, the points $(\sigma_n, \tilde{\delta}_n)$ are close to the straight line $\sigma_n = \tilde{\delta}_n$. Actually for values smaller than 1.6, nothing is caught by the Gaussian neighbourhood for this estimation, whereas for greater values, the estimation depends more on the signal than on the variability. The same remarks can be made about the choice of the size of the uniform kernel that seems to be appropriate. When comparing Fig. 8 with both Figs. 6 and 7, it can be seen that the range of $\tilde{\lambda}_n$ is slightly higher

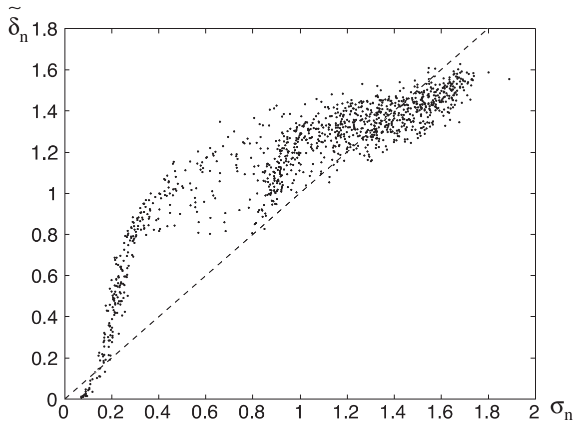

Fig. 7. Local variation measured by using a Gaussian kernel with a 1.6 standard deviation versus the statistical variation.

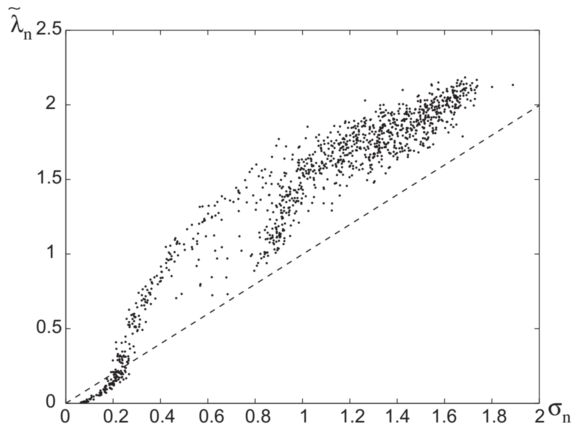

Fig. 8. Local variation measured by the length of the interval provided by the possibility distribution-based method versus the statistical variation.

Table 1
Correlation coefficients between the statistical standard deviation and the different measures of dispersion.

	$\gamma_{n,p}$	$\tilde{\gamma}_n$	$\delta_{n,p}$	$\tilde{\delta}_n$	$\lambda_{n,p}$	$\tilde{\lambda}_n$
Pearson	0.70	0.93	0.64	0.90	0.71	0.96
Spearman	0.64	0.92	0.63	0.90	0.67	0.95
Kendall	0.47	0.77	0.47	0.75	0.51	0.81

than the range of $\tilde{\gamma}_n$ and $\tilde{\delta}_n$. This is due to the fact that the measure $\tilde{\lambda}_n$ is just correlated to the noise level and is not an estimation of the standard deviation.

To objectively compare those three dispersion measures, we compute three correlation coefficients: Pearson, Spearman and Kendall. As can be seen in Table 1, the three averaged variability measures $\tilde{\gamma}_n$, $\tilde{\delta}_n$ and $\tilde{\lambda}_n$ are highly correlated with σ_n. The correlations between σ_n and the variability measures $\gamma_{n,p}$, $\delta_{n,p}$ and $\lambda_{n,p}$ are lower but are sufficient to show a dependency between these measures and the statistical variations of the set of images. We can notice that $\lambda_{n,p}$ is always more correlated with σ_n than the other variability measures $\gamma_{n,p}$ and $\delta_{n,p}$. The same remark is also true for $\tilde{\gamma}_n$, $\tilde{\delta}_n$ and $\tilde{\lambda}_n$. We can conclude that, in this experiment, the possibilistic approach that we propose seems to better quantify the noise level than the usual local approach.

5. Conclusion

In this article, we have presented a method for quantifying the noise level at each location of a signal. This method is based on replacing the conventional probabilistic by a possibilistic filtering approach. One of the main advantages of this method is the fact that nothing has to be assumed on the nature of the noise except its local ergodicity. Moreover, we put this possibilistic filtering approach in a more general framework of possibilistic signal processing. In this article, we do not show that the possibilistic layer, placed on the usual kernel-based signal processing that we propose, allows dealing with the noise all along a sequence of different kernel-based algorithms. This article is a first attempt to justify that the noise is easily handled in an isolated possibilistic signal processing method. Actually, the noise is handled by the possibilistic operation itself, which is an advantage compared to usual kernel-based approaches, where the noise estimation requires parallel computation. However, for more complicated procedures like sequences of kernel-based signal procedures, further research are required.

Acknowledgment

The authors thank Dr. Mariano-Goulard for providing them the data used in the experiments.

References

[1] S. Baker, I. Matthews, Lucas-Kanade 20 years on: a unifying framework, International Journal on Computer Vision 56 (3) (2004) 221–255.
[2] I. Bloch, H. Maitre, Fusion in image processing, in: I. Bloch (Ed.), Information Fusion in Signal and Image Processing, Wiley, 2008.
[3] J.P. Chilès, P. Delfiner, Geostatistics (Modeling Spatial Uncertainty), Wiley, New York, USA, 1999.
[4] B.R. Corner, R.M. Narayanan, S.E. Reichenbach, Noise estimation in remote sensing imagery using data masking, International Journal of Remote Sensing 24 (2003) 689–702.
[5] G. de Cooman, Integration and conditioning in numerical possibility theory, Annals of Mathematics and Artificial Intelligence 32 (2001) 87–123.
[6] D. Denneberg, Non-Additive Measure and Integral, Kluwer Academic Publishers, 1994.
[7] D. Dubois, Possibility theory and statistical reasoning, Computational Statistics and Data Analysis 51 (2006) 47–69.
[8] D. Dubois, H. Prade, Possibility Theory: An Approach to Computerized Processing of Uncertainty, Plenum Press, 1988.
[9] D. Dubois, H. Prade, S. Sandri, On possibility/probability transformations, in: Proceedings of Fourth IFSA Conference, Kluwer Academic Publishers, 1993, pp. 103–112.
[10] D. Dubois, L. Foulloy, G. Mauris, H. Prade, Probability–possibility transformations, triangular fuzzy sets, and probabilistic inequalities, Reliable Computing 10 (2004) 273–297.
[11] W.T. Freeman, E.C. Pasztor, O.T. Carmichael, Learning low-level vision, International Journal on Computer Vision 40 (1) (2000) 25–47.
[12] P. Goovaerts, Geostatistics for Natural Resources Evaluation, Oxford University Press, New York, 1997.
[13] G.E. Healey, R. Kondepudy, Radiometric CCD camera calibration and noise estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (1994) 267–276.
[14] J. Hutchinson, Culture, communication, and an information age madonna, IEEE Professional Communication Society Newsletter 45 (2001) 1–7.
[15] F. Jacquey, K. Loquin, F. Comby, O. Strauss, Nonadditive approach for gradient-based edge detection, in: ICIP'07, International Conference on Image Processing, San Antonio, Texas, USA, September, 2007, pp. 16-19.
[16] C. Liu, W.T. Freeman, R. Szeliski, S.B. Kang, Noise estimation from a single image, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006.
[17] K. Loquin, O. Strauss, On the granularity of summative kernels, Fuzzy Sets and Systems 159 (2008) 1952–1972.
[18] K. Loquin and O. Strauss, Imprecise functional estimation: the cumulative distribution case, in: SMPS08, Soft Methods in Probability and Statistic, Toulouse, France, 2008.
[19] D. Lowe, Object recognition from local scale-invariant features, Proceedings of the IEEE International Conference Computer Vision (1999) 1150–1157.
[20] C. Manders and S. Mann, Digital camera sensor noise estimation from different illuminations of identical subject matter, in: Proceedings of the Fifth International Conference on Information, Communications and Signal Processing, 2005, pp. 1292-1296.
[21] B.P. Marchant, R.M. Lark, Robust estimation of the variogram by residual maximum likelihood, Geoderma 140 (2007) 62–72.
[22] J.L. Marichal, Aggregation Operations for Multicriteria Decision Aid, Ph.D. Report, Department of Mathematics, University of Liège, 1998.
[23] G. Mauris, Inferring a possibility distribution from very few measurements, in: SMPS08, Soft Methods in Probability and Statistic, Toulouse, France, 2008.
[24] S.I. Olsen, Estimation of noise in images: an evaluation, CVGIP: Graphical Models and Image Processing 55 (1993) 319–323.
[25] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (7) (1990) 629–639.
[26] J. Portilla, V. Strela, M.J. Wainwright, E.P. Simoncelli, Image denoising using scale mixtures of Gaussians in the wavelet domain, IEEE Transactions on Image Processing 12 (11) (2003) 1338–1351.
[27] R.M. Rangayyan, M. Ciuc, F. Faghih, Adaptive-neighbourhood filtering of images corrupted by signal-dependent noise, Applied Optics 37 (1998) 4477–4487.

[28] M. Rombaut, Fusion in robotics, in: I. Bloch (Ed.), Information Fusion in Signal and Image Processing, Wiley, 2008.
[29] D. Schmeidler, Integral representation without additivity, Proceedings of the American Mathematical Society 97 (1986) 255–261.
[30] J.S. Simonoff, Smoothing Methods in Statistics, Springer, 1996.
[31] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, 1991.
[32] P. Walley, Towards a unified theory of imprecise probability, International Journal of Approximate Reasoning 24 (2000) 125–148.
[33] R. Zhang, P. Tsai, J. Cryer, M. Shah, Shape from shading: a survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (8) (1999) 690–706.