Noise quantization via possibilistic filtering

Kevin Loquin
IRIT
Université Paul Sabatier
118 Route de Narbonne
F-31062 Toulouse Cedex 9
Kevin.Loquin@irit.fr

Olivier Strauss
LIRMM
Université Montpellier II
161, rue Ada
F-34392 Montpellier Cedex 5
Olivier.Strauss@lirmm.fr

Abstract

In this paper, we propose a novel approach for quantifying the noise level at each location of a digital signal. This method is based on replacing the conventional kernel-based approach extensively used in signal filtering 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 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 advantageous over other methods since no assumption about the nature of the noise has to be made, except its local ergodicity.

Keywords. Signal processing, kernel methods, possibility distribution, noise quantization, Choquet integral.

1 Introduction

The reliability of a great number of signal processing methods inherently relies on the possibility of adjusting their parameters to account for noise level over the input signal. Examples of such procedures are image restoration, edge detection [18], motion estimation [1], denoising [26, 27], super-resolution [14], shape-from-shading [34], sensor fusion [3, 29] and feature extraction or segmentation [22].

Noise in a signal is usually referred to random variations of the measured signal. 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.

Noise is generally considered as being independent from the signal level and added to it. One of the most widely encountered model 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 signal's value 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).

In early approaches (see e.g [25]), noise estimation consisted in assuming stationarity of the random variations of the signal. The computation of the standard deviation of the noise were 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 a signal variation is due to the noise or to the signal itself, which can involve significant variations.

In more recent papers, some authors propose to abandon either stationarity or additivity of the noise. Rangayyan et al. [28] 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 [5].

Unfortunately, neither additive nor multiplicative random noise are good models for real signal contamination, even for instance, for conventional CCD sensor [18]. Therefore, many approaches [18, 16, 23] propose to model the acquisition noise as being Poisson distributed.

In these model-based approaches, the 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 type of the acquisition machine to be known.

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 [20] for direct or iterative, linear or non-linear algorithms and for filtering (stochastic, band pass, anti-aliasing, ...), geometrical transformations (rescaling, rotations, homographies, anamorphosis, ...), sampling rate conversion, fusion, for enhancing or removing details, etc. The kernels usually encountered are probability distributions: they are positive functions whose total weight (their integral in the infinite domain and their sum in the finite domain) sums to 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 take into account a lack of knowledge on the proper kernel to be used [21]. Indeed, a possibility distribution represents a convex hull of probability distributions and hence of kernels. In this 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 [17], 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 result of a possibilistic filtering on a signal and the input signal random noise. Actually, we discuss the fact that this approach is, to our opinion, in its spirit, better founded than the usual noise level estimators. Furthermore, we propose to highlight the empirical correlation between the length of the output of the interval-valued filtering and the input signal random noise on repeated acquisitions of real and synthetic images.

The paper is organized as follows. In section 2, we present how the digital filtering is performed by means of convolution kernels with unitary gains. We present the possibility distribution-based filtering, which is theoretically justified by Theorem 1. In section 3, we describe our method for estimating the noise level at each sample location of a signal. 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 convolution kernel-based linear filtering

2.1 Convolution kernel-based signal filtering

Let $S = (S_i)_{i=1,...,N}$ be a digital signal defined on N locations $\{\omega_1, ..., \omega_N\}$ of an underlying infinite domain Ω. Note that the locations $\{\omega_1, ..., \omega_N\}$ can be identified with their indices $\{1, ..., N\}$. Processing S by a filter, defined by its impulse response κ, mathematically corresponds to the discrete convolution of S by κ. This is why κ can also be called a convolution kernel. The value $\hat{S}_n$ of the filtered signal at the n^{th} location of $\{1, ..., N\}$ is thus obtained by:

$$\hat{S}_n = \sum_{i=1}^{N} S_i \kappa_{n-i}.$$

$\kappa_{n-\bullet} = (\kappa_{n-i})_{i=1,...,N}$ is the convolution kernel κ shifted to the location n of $\{1, ..., N\}$. We propose to denote this particular shifted kernel by $\kappa^n = (\kappa_i^n)_{i=1,...,N}$. $\hat{S}_n$ is thus obtained by:

$$\hat{S}_n = \sum_{i=1}^{N} S_i \kappa_i^n. \tag{1}$$

In many applications like low-pass filtering, the used convolution kernels are positive and have a unitary gain, i.e.

$$\sum_{i=1}^{N} \kappa_i = 1.$$

In that case, the convolution kernel can be seen as a probability distribution that induces a discrete probability measure P_κ, computed in this way:

$$\forall A \subseteq \Theta,\ P_\kappa(A) = \sum_{i \in A} \kappa_i.$$

For each location n, its associated shifted convolution kernel κ^n is still a probability distribution. Thus, expression (1) is equivalent to computing the expected value $\hat{S}_n$ of the signal S at the location n, considering the probability measure P_{κ^n} on $\{1, ..., N\}$, i.e.:

$$\hat{S}_n = \mathbb{E}_{P_{\kappa^n}}(S). \tag{2}$$

In that case, the filtered value of the signal can be interpreted as the expected value of the signal, knowing that the uncertainty concerning the location is modelled by P_{κ^n}. This interpretation is not very relevant because the aim of the filtering is not to try to evaluate the real value of a signal under uncertainty. The aim of the filtering is to modify the input signal according to the practitioner's needs. The only reason why we propose to rewrite the linear filtering with the expectation operator is that it enables us to deal with a family of convolution kernels by switching from the usual probability theory to imprecise probability theory.

2.2 Extension of signal filtering to possibility theory

By writing the linear filtering with a unitary gain filter as an expectation according to a probability measure, we open new perspectives to this approach by repositioning it in the field of new uncertainty theories. Instead of using an additive measure for each neighbourhood of a sample location, i.e. a probability measure, we propose to use the simple non-additive confidence measure called a possibility measure [9]. We propose to use this theory among others because of its computational simplicity. First, the possibility distribution is a tool that can be simply modelled by just a set of N weights on the locations $\{1, ..., N\}$, whereas most of the other imprecise probability theories [33] require more assessments. Besides, we propose to use the Choquet integral, that extends the usual linear expectation operator, by extending the convolution operator to possibility measures in place of probability measures. This tool is well know and very simply computed.

This section presents and interprets this new filtering approach, based on possibility measures and Choquet integrals, that enables a signal to be filtered by means of a family of convolution kernels.

2.2.1 A possibility distribution is a family of filters

A possibility measure is non-additive and possesses a dual confidence measure, called a necessity measure, denoted by N and computed in this way:

$$\forall A \subseteq \Theta,\ N(A) = 1 - \Pi(A^c). \tag{3}$$

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

$$\mathcal{M}(\Pi) = \{P \mid \forall A \subseteq \Theta, N(A) \leq P(A) \leq \Pi(A)\}.$$

This encoding property is due to the sensitivity analysis interpretation [32] of possibility theory.

A possibility measure can be defined from a possibility distribution π^n. Such a distribution is normalized in the sense that

$$\max_{i \in \Theta} \pi_i^n = 1.$$

Its associated possibility measure is obtained by:

$$\forall A \subseteq \Theta,\ \Pi_{\pi^n}(A) = \max_{i \in A} \pi_i^n.$$

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

$$\mathcal{M}(\pi^n) = \{\kappa^n \mid \forall A \subseteq \Theta, N_{\pi^n}(A) \leq P_{\kappa^n}(A) \leq \Pi_{\pi^n}(A)\}.$$

This family of convolution kernels being defined, the extension of the convolution (or expectation) operator has to be studied.

2.2.2 The possibilistic extension of the linear filtering

Since a possibility measure is non-additive, the conventional expectation operator cannot be used for filtering. The expectation operator must be replaced by its generalization, called the Choquet integral [6]. Using a Choquet integral and a possibility distribution leads to an interval-valued expectation, instead of a single value, whose upper and lower bounds are given by:

$$\overline{S}_n = \mathbb{C}_{\Pi_{\pi^n}}(S), \tag{4}$$

$$\underline{S}_n = \mathbb{C}_{N_{\pi^n}}(S). \tag{5}$$

The Choquet integral can be considered as a generalization of the conventional expectation operator since, when the used confidence measure is a probability measure, expressions (4) and (5) coincide and are equal to the conventional expectation operator (2).

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.

As a preliminary to the theorem (and its proof) justifying this assertion, some notations are necessary. Let us denote by $\mathcal{L}(\{1, ..., N\})$ the set of bounded sequences of weights on $\{1, ..., N\}$, i.e. $\forall I = (I_i)_{i=1,...,N} \in \mathcal{L}(\{1, ..., N\})$, $\max_{i=1,...,N} |I_i| < \infty$. In [32], this set is called the set of bounded gambles on $\{1, ..., N\}$. Denote $\mathcal{B}(\{1, ..., N\})$, the set of binary (i.e. $\{0, 1\}$-valued) sequences of weights on $\{1, ..., N\}$. Obviously, $\mathcal{B}(\{1, ..., N\}) \subset \mathcal{L}(\{1, ..., N\})$. $\mathcal{B}(\{1, ..., N\})$ can be seen as the set of events on $\{1, ..., N\}$.

Theorem 1. *Let π^n be a possibility distribution. $\forall S \in \mathcal{L}(\{1, ..., N\})$, $\forall \kappa^n \in \mathcal{M}(\pi^n)$,*

$$\mathbb{C}_{N_{\pi^n}}(S) \leq \mathbb{E}_{P_{\kappa^n}}(S) \leq \mathbb{C}_{\Pi_{\pi^n}}(S). \tag{6}$$

Moreover, the bounds are reached: $\forall S \in \mathcal{L}(\{1, ..., N\})$, $\exists \kappa_1^n, \kappa_2^n \in \mathcal{M}(\pi^n)$, such that

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

Proof. The natural extension principle [32] is required to prove Theorem 1. Note that the natural extension of a probability measure P, defined for all the events A of $\mathcal{B}(\{1, ..., N\})$, is the expectation according to P, defined for all S of $\mathcal{L}(\{1, ..., N\})$. Similarly, the natural extension of a possibility measure Π, defined for all the events A of $\mathcal{B}(\{1, ..., N\})$, is the Choquet integral with respect to Π, defined for all S of $\mathcal{L}(\{1, ..., N\})$[1].

[1] 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 $\mathcal{M}(\pi^n)$, noted $E(\mathcal{M}(\pi^n))$, is the same as the family of expectations dominated by the Choquet integral according to π^n, noted $\mathcal{M}(\mathbb{C}_{\Pi_{\pi^n}})$. This property of the natural extension can be found in Walley's book [32] for an upper prevision $\underline{P}$ and its associated set of linear previsions $\mathcal{M}(\underline{P})$. It is enough to conclude that $\forall S \in \mathcal{L}(\{1, ..., N\}), \forall \kappa^n \in \mathcal{M}(\pi^n)$,

$$\mathbb{C}_{N_{\pi^n}}(S) = \min\{\mathbb{E}_{P_{\kappa^n}}(S) : \kappa^n \in \mathcal{M}(\pi^n)\},$$
$$\mathbb{C}_{\Pi_{\pi^n}}(S) = \max\{\mathbb{E}_{P_{\kappa^n}}(S) : \kappa^n \in \mathcal{M}(\pi^n)\}.$$

□

This theorem is also valid for infinite domains. The proof is derived from domination theorems proved by Denneberg [7], proposition 10.3 and Schmeidler [30], proposition 3.

This propagation of the imprecision in the choice of the possibility distribution representing a family of kernels to the result of this new possibilistic filtering operation is very interesting. Using a possibility distribution allows the modelling of a lack of knowledge on the proper convolution kernel to be used. Using the generalized expectation operator (4) and (5) directly impacts this ill-knowledge on the output.

Note that in the case of a positive 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 explicitely computed by :

$$\overline{S}_n = \mathbb{C}_{\Pi_{\pi^n}}(S) = \sum_{i=1}^{N} \Pi_{\pi^n}(A_{(i)})(S_{(i)} - S_{(i-1)}), \quad (7)$$

$$\underline{S}_n = \mathbb{C}_{N_{\pi^n}}(S) = \sum_{i=1}^{N} N_{\pi^n}(A_{(i)})(S_{(i)} - S_{(i-1)}). \quad (8)$$

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

2.2.3 How to choose the possibilistic filter?

The use of a possibility distribution as a family of linear filters 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. Two hints for helping him to choose a possibility distribution are explored and provided in this paragraph.

First, we propose to use the triangular possibility distribution since it encodes (among others) all the symmetric convolution kernels with the same support [13]. Indeed, many algorithms (for example low-pass filtering) extensively use symmetric convolution kernels.

Second, probability/possibility transformations studied by Dubois et al. [13] can be used, when the practitioner has a vague idea of the convolution kernel to be used. The possibility distributions obtained by these transformations form families of convolution kernels including the kernel to approximate [10, 8]. The objective transformation results in the smallest family containing the original kernel and the subjective transformation [11, 12] results in a larger family of convolution kernels. The latter transformation should be preferred in case of little confidence in the choice of the original convolution kernel.

3 Noise estimation

3.1 Nuggets effect and local estimation by neighbourhood

Geostatistic is the branch of applied statistics that concentrates on the description of spatial patterns [4, 24, 15]. 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 increment of a particular observed property, function of the distance, between different sample locations, this increment[2] does not seem to vanish when the distance tends to 0. This discontinuity, which is supposedly due to geostatical noise, is called the "nuggets 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 signal increases with the radius Δ of the subset A. On the contrary, the

[2] Generally, the curve of the halve squared increments is plotted. This curve is called the sample variogram [4]

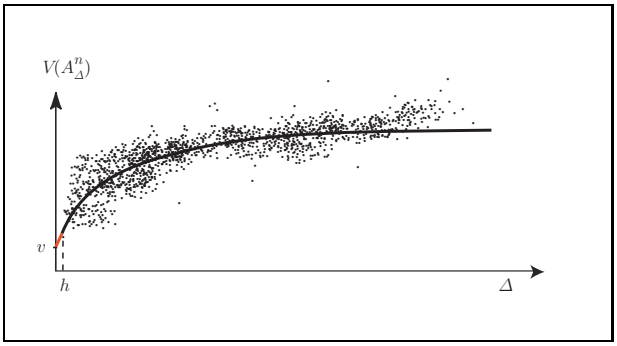

Figure 1: Qualitative example of variogram.

variability due to the measurement error 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^n_Δ is a neighbourhood of radius Δ of the n^{th} location ω_n, $V(A^n_\Delta)$, the variability of A^n_Δ is such that :

$$\lim_{\Delta \to 0} V(A^n_\Delta) = v_n, \tag{9}$$

with v_n being the variability due to measurement error at location n. This limit is known as the nuggets effect in the geostatistic field [15]. However, due to sampling, v_n 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 variability is to plot a variogram, i.e. to plot the variability of all the sampling locations of $n \in \{1, ..., N\}$, $V(A^n_\Delta)$, as a function of Δ. A manual fitting[3] is generally performed to provide an estimation of the nuggets effect, which is the value of the regression equation for a radius $\Delta = 0$. This estimation is denoted by v.

However, this method presupposes that the error is stationary all over the signal. Moreover, the choice to be made for a particular variogram equation 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.

A more pointwise estimation of these measurement errors can be obtained by means of a small neighbourhood around each sampled location. This approach is based on assuming local ergodicity. Local ergodicity states that the local variability of the signal in a small neighbourhood of a sampling point reflects the statistical variations of the signal at this location, due to measurement errors. The neighbourhood commonly used is a probability distribution defined over the set of pixels by $\kappa^n = (\kappa^n_i)_{i=1,...,N}$.

[3]Sometimes, automatic fitting procedures (which are not recommanded by geostatisticians), as regression analysis, are performed

The local variability computation leads to a weighted sum due to the additivity of the probability measure. Estimations of the nuggets effect are given by:

$$v_n = \sqrt{\sum_{k=1}^{N} (S_k - \hat{S}_n)^2 \kappa^n_k}, \tag{10}$$

if variability is measured by the standard deviation. And:

$$v_n = \sum_{k=1}^{N} |S_k - \hat{S}_n| \kappa^n_k, \tag{11}$$

if variability is measured by the mean error.

Most of the kernels used to perform this estimation are unimodal, centered and symmetric around the sample location n.

3.2 Noise quantization via possibilistic filtering

Our approach is also based on the assumption of local ergodicity. On top of that, it exploits the domination properties presented in section 2, i.e. of the fact that a possibility distribution can be seen as a family of convolution kernels.

Suppose you want to low-pass filter a signal with two different filters having the same cutoff frequency f_c. Such a filter eliminates from the input signal its component with a frequency higher than the cutoff frequency f_c (this is the explanation for the origin of 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 we apply this same filtering procedure to an input signal having frequencies beyond f_c. Then, generally, the output signals will be different, depending on the shape of the convolution kernel.

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[4], 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.

As mentioned before, the impulse responses of the usual linear low-pass filters are convolution kernels (uniform, Gaussian filters...). Since a possibility distribution is equivalent to a family of 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.

[4]This is the hypothesis underlying the low-pass filters

The imprecision or the dispersion in the result of a possibility distribution-based filtering is quantified by the length of the interval $[\underline{S}_n, \overline{S}_n]$, as defined by expressions (8) and (7).

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

$$\lambda_n = \overline{S}_n - \underline{S}_n. \tag{12}$$

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

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

$$\pi_{3\times 3} = \begin{pmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{pmatrix} \tag{13}$$

In the case of 1D signal processing, the used triangular neighbourhood of each sample location can be simply represented by the vector:

$$\pi_3 = \begin{pmatrix} 0.5 \\ 1 \\ 0.5 \end{pmatrix} \tag{14}$$

In order to weaken the influence of the signal variations on the noise level estimator that we propose, we have to choose the smallest possible neighbourhood. Under a π_3 or a $\pi_{3\times 3}$ possibilistic neighbourhood, is only the Kronecker possibility distribution that is equal to 1 on the estimation's location that would have led to a canonical estimation of S_n on the location n. This is why we propose to use π_3 or $\pi_{3\times 3}$ to estimate the noise level.

We conjecture that the length of the interval-valued estimate $[\underline{S}_n, \overline{S}_n]$ obtained with $\pi_{3\times 3}$ or with π_3 is an estimate of the noise level at the location n. This conjecture is illustrated by the experiments in section 4.

4 Experiments

4.1 Simulated noise experiment

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. With this set of

Figure 2: images of Lena with simulated Gaussian noise with standard deviations of 0, 30 and 60.

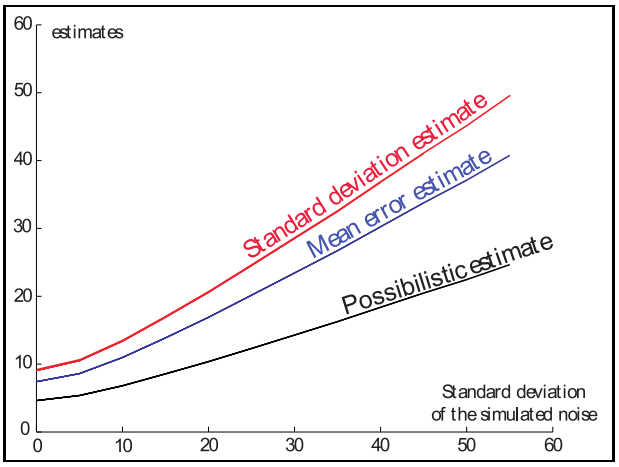

Figure 3: Usual and possibilistic local estimates of the noise level.

noisy images, we can directly compare the noise level estimates presented in this paper (10), (11) and (12) 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 Gaussian noise.

The average over all the pixels of the noisy images of the noise level estimates (10), (11) and (12) is plotted on Figure 3 versus the level of the simulated added noise. The highest curve corresponds to the standard deviation estimate, i.e. expression (10) with a 3×3 convolution kernel, the curve in the middle, corresponds to the mean error estimate, i.e. expression (11) with a 3×3 convolution kernel and the lowest curve corresponds to the possibility distribution-based noise level estimate, i.e. expression (12).

As can be seen on Figure 3, all these estimators are good markers of the noise level, since the three plotted curves are linear functions of the noise level. The part of the curves with small simulated noise levels (i.e. with stan-

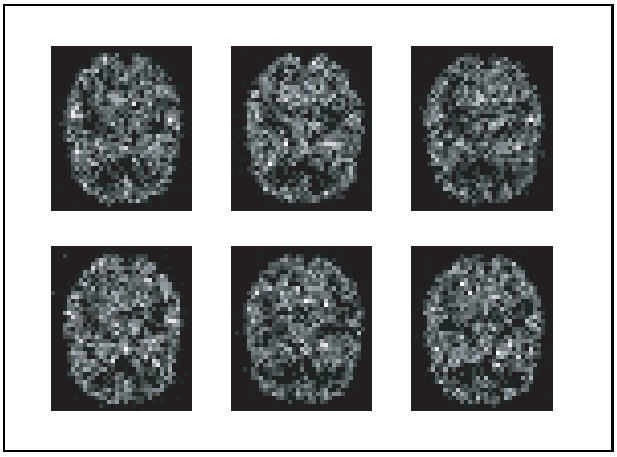

Figure 4: six images of the 1000 HBP direct acquisitions.

dard 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 can not 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 Real noise experiment

A Hoffman $2D$ brain phantom (Data Spectrum Corporation) 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 second; 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 n^{th} pixel within the p^{th} acquired image. Note that Figure 4 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 (12) is more correlated to the statistical variations of the image than the standard deviation noise estimation approach.

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 n^{th} pixel can be estimated by its standard deviation σ_n all over its different realizations:

$$\sigma_n = \sqrt{\frac{1}{999}\sum_{p=1}^{1000}(I_{n,p} - m_n)^2}, \tag{15}$$

with m_n, the weighted mean of the image at the n^{th} pixel:

$$m_n = \frac{1}{1000}\sum_{p=1}^{1000} I_{n,p}. \tag{16}$$

On the other hand, the local variation of the measurement in the neighbourhood of the n^{th} pixel within the p^{th} image can be estimated by computing the standard deviation via the expression (10) with a highly specific kernel (the same experiment made with expression (11) 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 [20, 31].

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 subsection 3.2. The local variation in the neighbourhood of the n^{th} pixel within the p^{th} image is estimated by the length $\lambda_{n,p}$ of each interval:

$$\lambda_{n,p} = \overline{I}_{n,p} - \underline{I}_{n,p}. \tag{17}$$

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 $\tilde{\lambda}_n = \frac{1}{1000}\sum_{p=1}^{1000}\lambda_{n,p}$.

Figure 5 plots $\tilde{\gamma}_n$ versus σ_n, as well as the straight line of equation $\sigma_n = \tilde{\gamma}_n$, figure 6 plots $\tilde{\delta}_n$ versus σ_n, as well as the straight line of equation $\sigma_n = \tilde{\delta}_n$ and figure 7 plots $\tilde{\lambda}_n$ versus σ_n, as well as the straight line of equation $\sigma_n = \tilde{\lambda}_n$.

These figures clearly show that all these estimations are, on average, correlated with σ_n. The choice of the value

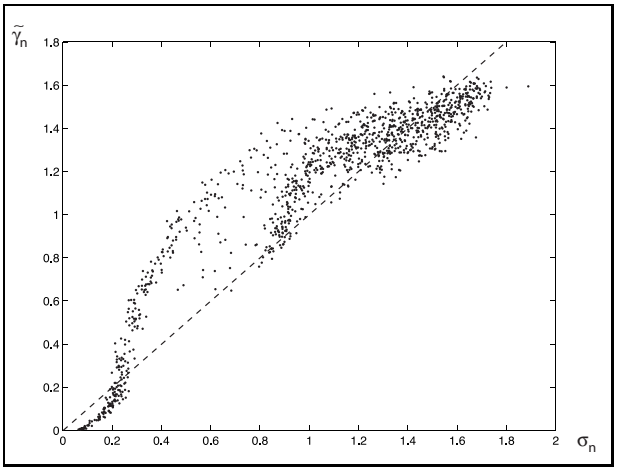

Figure 5: local variation measured by using a 3×3 uniform kernel versus the statistical variation.

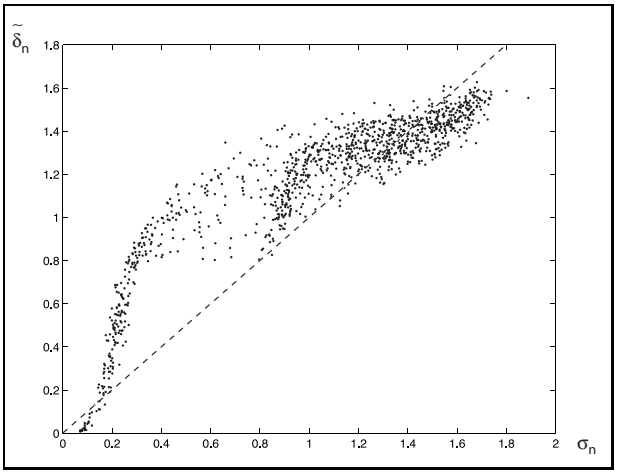

Figure 6: local variation measured by using a Gaussian kernel with a 1.6 standard deviation versus the statistical variation.

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 Figure 7 with both Figure 5 and 6, it can be seen that the range of $\tilde{\lambda}_n$ is slightly higher 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.

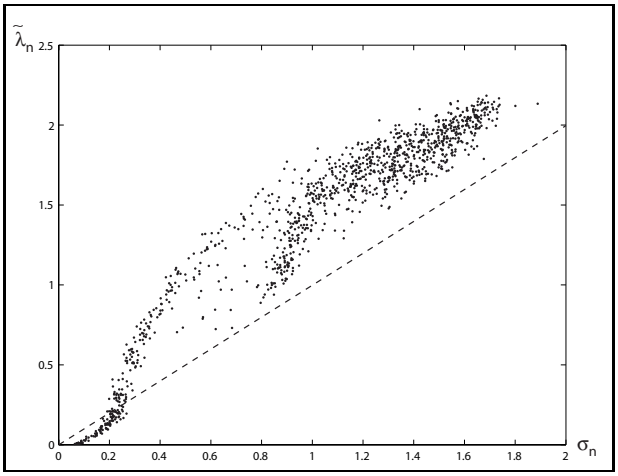

Figure 7: local variation measured by the length of the interval provided by the possibility distribution based method versus the statistical variation.

	$\gamma_{n,p}$	$\tilde{\gamma}_n$	$\delta_{n,p}$	$\tilde{\delta}_n$	$\lambda_{n,p}$	$\tilde{\lambda}_n$
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

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

As we conjecture that $\lambda_{n,p}$ could be regarded as a spread factor measuring the local noise level, we expect that two intervals $[\underline{I}_{n,p}, \overline{I}_{n,p}]$ and $[\underline{I}_{n,q}, \overline{I}_{n,q}]$ intersect for most of pairs (p, q) of images. We propose to complete this experimentation, by computing, for each pixel n, the ratio of the intervals that intersect versus the total number of tested intervals. We compute the same ratio using $\gamma_{n,p}$ and $\delta_{n,p}$ considered as spread factors measuring statistical standard deviations: we then test each couple of intervals $[I_{n,p} - 3\gamma_{n,p}, I_{n,p} + 3\gamma_{n,p}]$ and $[I_{n,p} - 3\delta_{n,p}, I_{n,p} + 3\delta_{n,p}]$. Since the 3σ interval is usually assumed to be the 99% confidence interval, one can expect a high rate of overlapping. Table 2 presents the average ratio for all the pixels

	with all pixels	only with pixels such that $I_n > 3$
Uniform kernel	0.11	0.88
gaussian kernel	0.13	0.89
possibility distribution	0.98	0.92

Table 2: Ratio of intersecting confidence intervals.

of the image and for only the pixels with a value greater than three.

As can be seen easily on Table 2, the possibility distribution based confidence interval fulfils a 98% intersecting intervals while the usual probabilistic based confidence intervals are far from this 99% ratio. The bad ratio of the other methods is mainly due to the fact that the spread factor is underestimated by these methods for low values (as it can be easily seen on Figures 5 to 7). In fact, selecting only the pixels whose level always exceeds a certain level over the different realizations increases the score of the probabilistic based methods. In fact, by assuming that the measured values are Poisson distributed, a local Gaussian approximation can be valid except for small values of the illumination signal.

5 Conclusion

In this article, we have presented a method for quantifying the noise level at each sample location of a signal. This method is based on replacing the conventional probabilistic by a possibilistic filtering approach. One of the main advantage of this method is the fact that nothing has to be assumed on the nature of the noise except its local ergodicity. Moreover, when a possibilistic approach is used in signal processing, the noise estimation is propagated all along the different steps of the algorithm by the model itself, which is an advantage compared to usual kernel based approaches, where the noise estimation requires a parallel computation.

6 Acknowledgment

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

References

[1] S. Baker and I. Matthews, Lucas-Kanade 20 years on: a unifying framework, *International Journal on Computer Vision*, vol. 56(3), 2004, pp. 221255.

[2] I. Bloch and H. Maitre, Fuzzy mathematical morphologies: A comparative study, *Pattern Recognition*, vol. 28, 1995, pp. 1341-1387.

[3] I. Bloch and H. Maitre, Fusion in image processing, In: *Information Fusion in Signal and Image Processing, I. Bloch (Ed.), Wiley*, 2008.

[4] J.P. Chilès et P. Delfiner, Geostatistics (Modeling Spatial Uncertainty), *Wiley, New-York, U.S.A.*, 1999.

[5] B.R. Corner, R.M. Narayanan, and S.E. Reichenbach, Noise estimation in remote sensing imagery using data masking, *International Journal of Remote Sensing*, vol. 24, 2003, pp. 689-702.

[6] G. de Cooman, Integration and conditioning in numerical possibility theory, *Annals of Mathematics and Artificial Intelligence*, vol. 32, 2001, pp. 87-123.

[7] D. Denneberg, Non-Additive Measure and Integral, *Kluwer Academic Publishers*, 1994.

[8] D. Dubois, Possibility theory and statistical reasoning, *Computational Statistics and Data Analysis*, vol. 51, 2006, pp. 47-69.

[9] D. Dubois and H. Prade, Possibility theory: an approach to computerized processing of uncertainty, *Plenum Press*, 1988.

[10] D. Dubois, H. Prade, and S. Sandri, On possibility/probability transformations, *Proceedings of Fourth IFSA Conference, Kluwer Academic Publ*, 1993, pp. 103-112.

[11] D. Dubois, H. Prade, and P. Smets, New semantics for quantitative possibility theory, *ISIPTA01, 2nd International Symposium on Imprecise Probabilities and Their Applications, Ithaca, New York, USA, June*, 2001.

[12] D. Dubois, H. Prade, and P. Smets, A definition of subjective possibility, *International Journal of Approximate Reasoning*, vol. 48, 2008, pp. 352-364.

[13] D. Dubois, L. Foulloy, G. Mauris, and H. Prade, Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities, *Reliable Computing*, vol. 10, 2004, pp. 273-297.

[14] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, Learning low-level vision, *International Journal on Computer Vision*, vol. 40(1), 2000, pp. 2547.

[15] P. Goovaerts, Geostatistics for natural resources evaluation, *Oxford University Press New York*, 1997.

[16] G.E. Healey and R. Kondepudy, Radiometric CCD camera calibration and noise estimation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, 1994, pp. 267-276.

[17] F. Jacquey, K. Loquin, F. Comby, and O. Strauss, Nonadditive approach for gradient-based edge detection, *ICIP07, Int. Conf. on Image Processing, San Antonio, Texas, USA, September*, 2007, pp. 16-19.

[18] C. Liu, W.T. Freeman, R. Szeliski, and S.B. Kang, Noise Estimation from a Single Image, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.

[19] K. Loquin, De l'utilisation des noyaux maxitifs en traitement de l'information, *PhD report, LIRMM, Université Montpellier II*, 2008

[20] K. Loquin and O. Strauss, On the granularity of summative kernels, *Fuzzy Sets and Systems*, vol. 159, 2008, pp. 1952-1972.

[21] K. Loquin, and O. Strauss, Imprecise functional estimation: the cumulative distribution case, *SMPS08, Soft Methods in Probability and Statistic*, Toulouse, 2008.

[22] D. Lowe, Object recognition from local scale-invariant features, *In Proc. IEEE Intl Conf. Computer Vision*, 1999, pp. 11501157.

[23] C. Manders and S. Mann, Digital Camera Sensor Noise Estimation from Different Illuminations of Identical Subject Matter, *Proceedings of the Fifth International Conference on Information, Communications and Signal Processing*, 2005, pp. 1292-1296.

[24] B.P. Marchant and R.M. Lark, Robust estimation of the variogram by residual maximum likelihood, *Geoderma*, vol. 140, 2007, pp. 62-72.

[25] S.I. Olsen, Estimation of noise in images: an evaluation, *CVGIP: Graphical Models and Image Processing*, vol. 55, 1993, pp. 319-323.

[26] P. Perona and J. Malik, Scale-space and edge detection using anisotropic diffusion, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12(7), 1990, pp. 629639.

[27] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, Image denoising using scale mixtures of Gaussians in the wavelet domain, *IEEE Transactions on Image Processing*, vol. 12(11), 2003, pp. 13381351.

[28] R.M. Rangayyan, M. Ciuc, and F. Faghih, Adaptive-neighborhood filtering of images corrupted by signal-dependent noise, *Applied Optics*, vol. 37, 1998, pp. 4477-4487.

[29] M. Rombaut, Fusion in robotics, In: *Information Fusion in Signal and Image Processing, I. Bloch (Ed.), Wiley*, 2008.

[30] D. Schmeidler, Integral representation without additivity, *Proceedings of the American Mathematical Society*, vol. 97, 1986, pp. 255261.

[31] J.S. Simonoff, Smoothing Methods in Statistics, *Springer*, 1996.

[32] P. Walley, Statistical reasoning with imprecise probabilities, *Chapman and Hall*, 1991.

[33] P. Walley, Towards a unified theory of imprecise probability, *International Journal of Approximate Reasoning*, vol. 24, 2000, pp. 125-148.

[34] R. Zhang, P. Tsai, J. Cryer, and M. Shah, Shape from shading: A survey, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21(8), 1999, pp.690706.