NON-ADDITIVE APPROACH FOR GRADIENT-BASED EDGE DETECTION

Florence JACQUEY, Kevin LOQUIN, Frédéric COMBY and Olivier STRAUSS

LIRMM, 161, rue Ada, 34392 Montpellier Cedex 5 France

ABSTRACT

In this paper, we propose a new method to perform the first derivative estimation of a discrete intensity distribution. This approach is based on a non-additive aggregation process and provides an estimate of the gradient as intervals instead of single values. These intervals are used to threshold a gradient-based edge detection and therefore discard spurious detections due to noise.

Index Terms— Edge detection, Choquet integral, capacity, summative and non-summative kernels, possibility, fuzzy numbers.

1. INTRODUCTION

Edges in images are areas with strong intensity contrasts. Gradient based edge detection consists of detecting maxima and minima in the image first derivative. However, edges are not the only features maximizing (or minimizing) this first derivative. Acquisition (or reconstruction) noise, digitalization and spurious local textures induce undesirable discontinuities in the intensity function.

Computation of the first derivative, in digital images, relies on the concept of fitting a continuous surface locally to the data. One of the most advanced approaches makes use of summative kernels whose role is to define a weighted neighborhood of each sampled location. This neighborhood provides an interplay between continuous and discrete. It also aims at using a hypothesized ergodicity to reduce the effect of random noise by providing regularization in the interpolation process.

Noise reduction and edge detection lead to antagonist criteria. The Canny operator was designed to be an optimal edge detector according to particular criteria. However there are other detectors around that also claim to be optimal with respect to slightly different criteria.

Estimation of the derivative of the continuous signal s resulting from the interpolation of discrete signal S via the summative kernel κ can easily be obtained by convolving S with the first derivative $d\kappa$ of kernel κ. Most of the usual approaches lead to a derivative kernel that can be obtained by subtracting two usual kernels:

$$d\kappa(x) = \eta^+(x + \epsilon_x) - \eta^-(x - \epsilon_x). \tag{1}$$

Even when a supposed optimal filter is used, noisy images produce spurious discontinuities that have to be removed by a tracking process involving an arbitrary threshold depending on the image noise. The threshold arbitrariness reduces the detection robustness. Indeed, the edge detection highly relies on the appropriateness of the hypothesized optimal threshold, especially with tomographic reconstructed images whose noise is usually assumed to be Poisson distributed and therefore not uniform on the whole image the detection robustness. In this paper, we present a method producing a natural threshold estimation, depending on local noise level, at each sampled location. The underlying idea of this work is the following: when estimating a piece of information from a set of noisy data with a kernel-based approach, variations due to measurement errors can be estimated by slightly varying the shape (but not the size) of the kernel. Such a variation can be simply achieved by replacing the conventional summative kernel approach by a non-summative kernel approach. We come up to a gradient estimator that produces intervals instead of single values. The length of each interval is highly correlated with the local noise level. A location is declared not being an edge if the value 0 belongs to the estimated interval.

2. SUMMATIVE AND NON-SUMMATIVE KERNEL

Summative or non-summative kernels are functions that can be thought as weighted neighborhoods.

Summative kernels are $\mathbb{R}$-valued functions κ defined on a domain Ω, satisfying the summative normalization property

$$\int_\Omega \kappa(\omega)d\omega = 1. \tag{2}$$

In image processing they are usually positive and unimodal as in the rest of this paper. Positivity implies that they can be seen as probability density functions, with whom we can associate, for any $A \subseteq \Omega$, a probability measure $P_\kappa(A) = \int_A \kappa(\omega)d\omega$.

Estimation of a given signal S in a given summative or probabilistic neighborhood κ is the expectation of this signal according to κ. That is:

$$\mathbb{E}_\kappa(S) = \int_\Omega SdP_\kappa = \int_\Omega S(\omega)\kappa(\omega)d\omega. \tag{3}$$

Non-summative kernels are $[0, 1]$-valued functions π defined on a domain Ω, satisfying the maximitive normalization

property

$$\max_{\omega \in \Omega} \pi(\omega) = 1. \qquad (4)$$

They can be seen as possibility distribution functions [1], with whom we can associate, for any $A \subseteq \Omega$, a possibility measure $\Pi_\pi(A) = \sup_{\omega \in A} \pi(\omega)$. A possibility measure is a special case of concave Choquet capacity ν. The conjugate ν^c of a concave capacity, defined by $\forall A \subseteq \Omega, \nu^c(A) = 1 - \nu(A^c)$ is a convex capacity. The conjugate N_π of a possibility measure Π_π is a necessity measure.

Those two kinds of neighborhood definitions (summative and non-summative) are linked by the fact that a concave capacity ν can encode a special family of probability measures (or equivalently additive measures), noted $core(\nu^c)$ and defined by

$$core(\nu^c) = \{P_\kappa \mid \forall A \subseteq \Omega, \nu^c(A) \leq P_\kappa(A) \leq \nu(A)\}. \qquad (5)$$

Now when the summative neighborhood is replaced by a capacity, estimation (3) is replaced by an estimation involving a Choquet integral [2, 3]

$$\mathbb{C}_\nu(S) = (C) \int_\Omega Sd\nu = \int_0^{+\infty} \nu(\{\omega : S(\omega) > \alpha\})d\alpha. \qquad (6)$$

The encoding property of a capacity ν (5) is transferred to the estimation as stressed by these inequalities:

$$\forall \kappa | P_\kappa \in core(\nu^c), \ \mathbb{C}_{\nu^c}(S) \leq \mathbb{E}_\kappa(S) \leq \mathbb{C}_\nu(S) \qquad (7)$$

See [4] proposition 3 and [2] proposition 10.3 for proofs.

As mentioned by Dubois and al. in [5], any summative kernel κ, which is centered and symmetric, verifies the domination property

$$\forall A \subseteq \Omega, P_\kappa(A) \leq \Pi_T(A), \qquad (8)$$

with T the triangular possibility distribution. This means that T encodes any centered symmetric summative kernel, which justifies our choice for the use of the triangular possibility distribution for the interplay between continuous and discrete. The domination of possibility measures can easily be extended to higher dimensions. For the dimension of 2, on a space $\Omega_1 \times \Omega_2$, a product of marginal possibility measures Π_1 and Π_2 is defined in [1] by:

$$\forall (A, B) \subseteq \Omega_1 \times \Omega_2, \Pi_{12}(A, B) = min(\Pi_1(A), \Pi_2(B)).$$

Suppose that marginally, $\Pi_1 \geq P_1$ and $\Pi_2 \geq P_2$, with P_1 and P_2 probability measures on, respectively, Ω_1 and Ω_2. Then Π_{12} dominates P_{12} defined by:

$$\forall (A, B) \subseteq \Omega_1 \times \Omega_2, P_{12}(A, B) = P_1(A)P_2(B).$$

$min(\Pi_1(A), \Pi_2(B)) \geq min(P_1(A), P_2(B)) \geq P_1(A)P_2(B)$, because $P \leq 1$. Therefore, by definitions of P_{12} and Π_{12}, $\Pi_{12}(A, B) \geq P_{12}(A, B)$.

3. DERIVATION WITH NON-SUMMATIVE KERNEL

As presented in section 1, most of the usual derivative kernels can be obtained by subtracting two summative kernels η^+ and η^-. Convolving signal S with the derivative kernel given by expression (1) is equivalent to summing up the convolution of S with the kernel η^+ and the convolution of the opposite of S with kernel η^-:

$$DS(x) = (\eta^+(x + \epsilon_x) - \eta^-(x - \epsilon_x)) * S(x)$$
$$= \eta^+(x + \epsilon_x) * S(x) + \eta^-(x - \epsilon_x) * (-S(x))$$

Therefore let $\kappa^+(x) = \eta^+(x + \epsilon_x)$ and $\kappa^-(x) = \eta^-(x - \epsilon_x)$. The estimate of $DS(x)$, derivative of S at location x, is given by:

$$DS(x) = \mathbb{E}_{\kappa^+}(S) + \mathbb{E}_{\kappa^-}(-S) \qquad (9)$$

If κ^+ (resp. κ^-) is dominated by a concave capacity ν^+ (resp. ν^-) then, the value $\mathbb{E}_{\kappa^+}(S)$ (resp. $\mathbb{E}_{\kappa^-}(-S)$) belongs to $[\mathbb{C}_{\nu^{+c}}(S), \mathbb{C}_{\nu^+}(S)]$ (resp. $[\mathbb{C}_{\nu^{-c}}(-S), \mathbb{C}_{\nu^-}(-S)]$). If κ^+ (resp. κ^-) is a bounded symmetric positive kernel, then the triangular possibility distribution can be used to ensure this domination (see section 2).
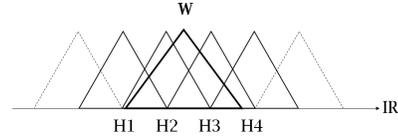
Fig. 1. Strong triangular partition of the real line and fuzzy neighborhood W.

The signal to be processed being discrete, a kernel has to ensure the interplay between continuous and discrete. In our approach, this interplay makes use of triangular fuzzy numbers. The continuous signal is supposed to be known on imprecise locations defined by the fuzzy partition $(H_i)_{i \in \{1, ..., N\}}$ depicted on Figure (1). We aim at transferring the discrete knowledge of the signal to any weighted neighborhood W. This transfer is ensured by using a Choquet integral. The signal being discrete, expression (6) can be simply computed by:

$$\mathbb{C}_{\nu_W}(S) = \sum_{n=1}^N S_{(n)}[\nu_W(A_{(n)}) - \nu_W(A_{(n+1)})], \qquad (10)$$

where S_n is the real positive value associated with the nth cell $\mathbf{H_n}$ and $(.)$ indicates a permutation such that $S_{(1)} \leq S_{(2)} \leq ... \leq S_{(N)}$. The $A_{(n)} = \{\mathbf{H_{(n)}}, ..., \mathbf{H_{(N)}}\}$ are binary coalitions of cells whose associated values are superior or equal to $S_{(n)}$.

Computation of $\mathbb{C}_{\nu_W^c}(S)$ can be simply achieved while noticing that $\mathbb{C}_{\nu_W^c}(S) = -\mathbb{C}_{\nu_W}(-S)$.

We propose to use the pignistic transfer to define the capacity used in Equation (10). Such a capacity is concave and its

expression is:

$$\nu_W(A_{(n)}) = \frac{|W \bigcap A_{(n)}|}{|W \bigcap A_{(1)}|}. \qquad (11)$$

Expression (9), associated with capacity ν_W, leads to an imprecise estimation of the first derivative of S:

$$[\underline{DS}, \overline{DS}] = [-\mathbb{C}_{\nu_W^+}(-S), \mathbb{C}_{\nu_W^+}(S)] + [-\mathbb{C}_{\nu_W^-}(S), \mathbb{C}_{\nu_W^-}(-S)]$$
$$= [-\mathbb{C}_{\nu_W^+}(-S) - \mathbb{C}_{\nu_W^-}(S), \mathbb{C}_{\nu_W^+}(S) + \mathbb{C}_{\nu_W^-}(-S)].$$

Since the capacity ν_W is concave, this interval contains all estimates of the derivative provided by a probability density function dominated by ν_W (i.e. all probability density functions defined by a kernel dominated by the possibility distribution defined by W). While the length of this interval increases with the noise level, this derivative can be supposed to be null if $0 \in [\underline{DS}, \overline{DS}]$.

4. EDGE DETECTION WITH NON-SUMMATIVE KERNEL BASED GRADIENT

The discrete image representation introduces an imprecision on the grey level localization. Let $P_{i,j}$ be the pixel located at (i, j) on the image. Its grey level localization is unknown within the 2D interval $[i - \Delta_h, i + \Delta_h] \times [j - \Delta_v, j + \Delta_v]$, where Δ_h (resp. Δ_v) is half the width of the horizontal (resp. vertical) sampling. This imprecision is due to the spatial sampling and can be modeled by a fuzzy partition [6]. Each pixel of the image is considered as a bi-dimensional imprecise quantity. The 1-D triangular fuzzy numbers presented in section 2 are extended in 2D pyramidal fuzzy numbers (Figure 2(a)) considering the t-norm min (cartesian product of two 1-D triangular fuzzy numbers), providing a strong fuzzy partition of the image.

(a) Fuzzy pixel   (b) 2D non-summative kernels
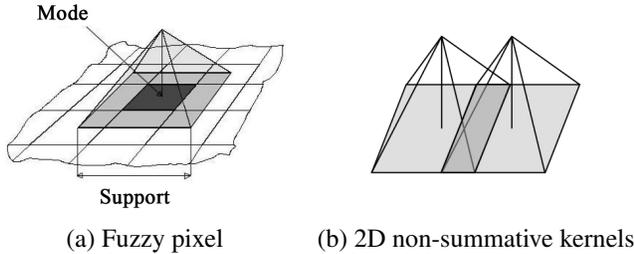
Fig. 2. Fuzzy representation.

As shown in section 2, the domination property still holds. Therefore, the properties presented in section 3 can be easily extended to provide an imprecise estimation of the gradient. We present this approach with the x-derivative. Let $\kappa(x, y) = \kappa_x(x)\kappa_y(y)$ be a separable kernel used to perform

the derivation, the 2D extension of expression (1) is given by:

$$\frac{\delta}{\delta x}(\kappa(x, y)) = \frac{\delta}{\delta x}(\kappa_x(x)\kappa_y(y)) = \frac{\delta}{\delta x}(\kappa_x(x))\kappa_y(y)$$
$$= (\eta^+(x + \epsilon_x) - \eta^-(x - \epsilon_x))\kappa_y(y).$$

Let $\kappa_x^+ = \eta^+(x + \epsilon_x)\kappa_y(y)$ and $\kappa_x^- = \eta^-(x - \epsilon_x)\kappa_y(y)$. Therefore, an estimate of the derivative of the 2D signal S along the x-axis (GS_x) can be given by simply adding the convolution of S with the summative kernel κ_x^+ to the convolution of $(-S)$ with the summative kernel κ_x^-:

$$GS_x = \mathbb{E}_{\kappa_x^+}(S) + \mathbb{E}_{\kappa_x^-}(-S). \qquad (12)$$

If ν_x^+ (resp. ν_x^-) is a concave capacity that dominates the kernel κ_x^+ (resp. κ_x^-) then the domination can be easily transferred to the estimation process by using the Choquet integral (10). Then, the imprecise estimation of GS_x is a simple 2D extension of the approach presented in section 3. It involves two pyramidal non-summative kernels depicted on Figure 2(b) dominating a family of kernels κ_x^+ and κ_x^-. Imprecise estimation of GS_y is achieved in the same way. Within this approach, the edge detection is performed on the median estimate of the gradient $[\underline{GS_x}, \overline{GS_x}] \times [\underline{GS_y}, \overline{GS_y}]$. An extremum of this function is discarded if

$$(0, 0) \in [\underline{GS_x}, \overline{GS_x}] \times [\underline{GS_y}, \overline{GS_y}].$$

5. EXPERIMENTAL RESULTS

In this section, we compare our approach with the Canny-Deriche [7], Shen-Castan [8] and Prewitt approaches. This comparison uses the artificial image depicted in Figure 3(a). This image has been corrupted with two kinds of noise. We have applied, first, a Gaussian noise with an increasing variance σ (Figure 3(b) $\sigma = 25$) to test the robustness of the edge detection to the noise level and, then, a Poisson noise to highlight the robustness of the method within variations of the noise level. This image mostly contains sharp transitions and therefore is perfectly appropriate for edge detection by using the conventional approaches.

(a) original   (b) Gaussian noise   (c) Poisson noise
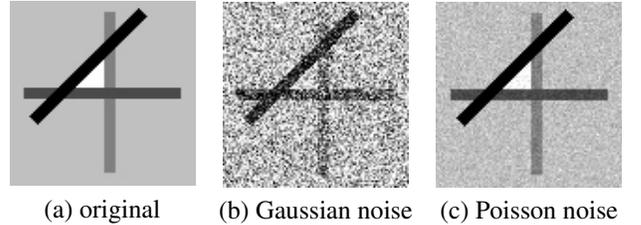
Fig. 3. Artificial image with Gaussian and Poisson noise.

The operators performances are quantified by the parameter P_1 defined by Fram and Deutsch in [9]. P_1 measures the sensitivity of the detector in presence of noise. $P_1 = 1$ if the detecting of edges is optimal.

In a first experiment, edges are detected without any threshold. Figure 4(a) shows that all approaches except ours fail when $\sigma > 3$. In a second approach, we have applied an arbitrary threshold chosen, for each method, to provide the best result within the P_1-criterion. Figure 4(b) shows that the Prewitt filter is the most sensible to noise. The Canny-Deriche and Shen-Castan approaches have a similar behavior: while $\sigma < 25$, optimal edges are correctly detected. The results decrease until $\sigma < 40$. After, edge and noise are confused. Within our approach, an optimal edge is detected while $\sigma < 30$. When $\sigma > 40$, all gradient estimates are discarded as being null.

(a) without threshold

(b) with an arbitrary threshold
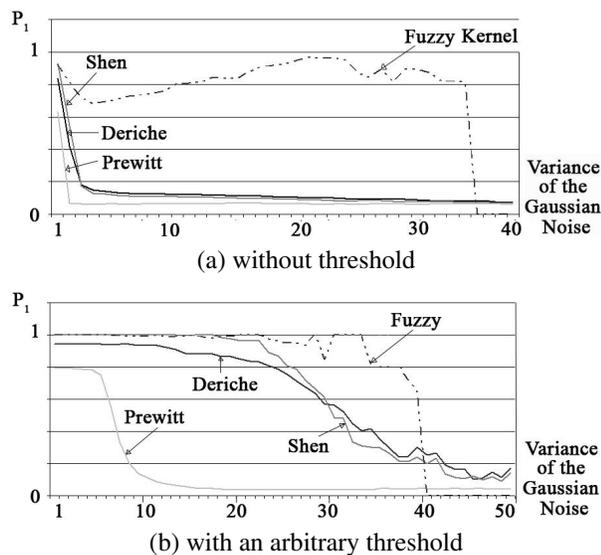
Fig. 4. Comparative P1 values on Gaussian noised image.

Approach	Prewitt	Deriche	Shen	Fuzzy
P_1	0.089676	0.14975	0.141	0.69881

Table 1. P_1 values on the Poisson noised image

When edge detection is applied on tomographic images, the threshold based approach usually fails because the noise level is not uniformly distributed in the image. The noise in reconstructed tomographic images is usually assumed to be Poisson distributed . In the last experiment, we have corrupted the test image with a Poisson noise. Results of this experiment are reported in Table 1. It shows that the fuzzy approach is 8 times less sensitive than the Prewitt approach and 5 times less than the Canny-Deriche and Shen-Castan approaches.

6. CONCLUSIONS

In this paper, we have presented a new approach for gradient-based edge detection. This approach involves non-summative

kernels and Choquet integrals and produces, as output, interval-valued estimation of the gradient. When used to detect edges on a noisy image, our approach, contrarily to classical methods, does not need an arbitrary threshold to discard spurious edges, while the threshold is naturally provided by the gradient estimation. Tests on synthetic images have illustrated the good performances of this approach with respect to state of art methods. Tests on real and particulary tomographic images have also been performed, but not reported here, while the appropriateness of the edge detection cannot be measured by an objective criterion, since the real edges of the image are usually unknown. These tests have shown the high reliability of the estimated noise level when repeating the experiment on different acquisitions of the same image.

7. REFERENCES

[1] D. Dubois and H. Prade, *Possibility Theory An Approach to Computerized Processing of Uncertainty*, Plenum Press, 1988. 2

[2] D. Denneberg, *Non-Additive Measure and Integral*, Kluwer Academic Publishers, Dordrecht, 1994. 2

[3] J.-L. Marichal, *Aggregation Operations for Multicriteria Decision Aid*, Ph.D. thesis, Department of Mathematics, University of Liège, Liège, Belgium, 1998. 2

[4] D. Schmeidler, "Integral representation without additivity," in *Proc. of the Am. Math. Soc.*, June 1986, vol. 97, pp. 255–261. 2

[5] D. Dubois, H. Prade, L. Foulloy, and G. Mauris, "Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities," *Reliable Computing*, vol. 10, pp. 273–297, 2004. 2

[6] Olivier Strauss and Frederic Comby, "Fuzzy morphology for omnidirectional images," in *ICIP*, Genova, Italy, September 11-14 2005, vol. 2, pp. 141–144. 3

[7] R. Deriche, "Fast algorithms for low-level vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 78–87, 1990. 3

[8] J. Shen and S. Castan, "An optimal linear operator for step edge detection computer vision," *Graphics, and Image Processing*, vol. 54, pp. 112–133, 1992. 3

[9] E. S. Deutsch and J. R. Fram, "A quantitative study of the orientational bias of some edge detector schemes," *IEEE Transactions on Computers*, March 1978. 3