Journal of Mathematical Imaging and Vision 16: 199–222, 2002

Multiresolution Design of Aperture Operators

ROBERTO HIRATA JUNIOR
SENAC College of Computer Science and Technology Rua Galvão Bueno, 430 - 01506-000 - São Paulo - SP - Brazil

MARCEL BRUN & JUNIOR BARRERA
Departamento de Ciência da Computação, Instituto de Matemática e Estatística - USP Rua do Matão, 1010 - 05508-900 - São Paulo - SP - Brazil

EDWARD R. DOUGHERTY
Department of Electrical Engineering, Texas A&M University College Station - 77843-3128 - TX - United States

Abstract. The design of an aperture operator is based on adequately constraining the spatial domain and the graylevel range in order to diminish the space of operators and, consequently, the estimation error. The design of a resolution constrained operator is based on adequately combining information from two or more different resolutions and has the same motivation, that is, diminish the space of operators to facilitate design. This paper joins these approaches and studies multiresolution design of aperture operators for grayscale images. Spatial resolution constraint, range resolution constraint and the combination of both constraints are characterized, and the error increase by using the constrained filter in place of the optimal unconstrained one is analyzed. Pyramidal multiresolution design involves applying the resolution constraint approach hierarchically, from the higher to the lower resolution space. These approaches are also characterized and their error increase analyzed. The system that has been implemented to design pyramidal multiresolution operators is described and has its complexity (memory and runtime) analyzed. Several simulations and two applications for deblurring are shown and compared to optimal linear filters. The results confirm the usefulness of the approach.

Keywords: multiresolution, nonlinear filter, optimal filter

1. Introduction

Optimal windowed filters [1–4] estimate a pixel value in an ideal image based on a window of values in an observed image. The filters are defined by functions over the random vector in the window. Since there is rarely a suitable model with which to perform optimization analytically, the standard design approach is to estimate the optimal filter from realizations of pairs of ideal and observed images. Estimation is problematic when the window is large owing to the exponentially growing demand for data to achieve good estimation precision [5]. The data requirement can be mitigated by constraining filter optimization to a subclass of all possible filters. Various constrained filter classes have been studied [6–22]. Each yields a suboptimal filter having increased error in comparison with the unconstrained optimal filter; however, each produces increased precision (decrease in estimation error). A constraint is beneficial if it results in designed filters having better expected performance. The worth of a constraint depends on the sample size. A constraint is typically beneficial for samples below a certain size, but not for samples beyond that size. Its worth also depends on the image class under consideration.

This paper joins two filter constraints. Aperture filters [7] have been used successfully for design of

grayscale filters, including deblurring. Multiresolution filters [10] have been used successfully for binary filtering and their benefit has been mathematically quantified in that context. Aperture filters form a class of nonlinear operators that filter grayscale images by operating in a finite window that is constrained both temporally (spatially) and in amplitude. These constraints allow a tractable design procedure to be applied by limiting the search space to a manageable size. Multiresolution design takes advantage of the following observation: optimally filtering an image at high resolution is better than optimally filtering at low resolution; however, increased resolution brings an increase in the number of random variables for a fixed size domain and a concomitant increase in error when estimating the optimal filter. More generally, a pyramidal multiresolution design can be employed that uses data from different resolutions.

Following this Introduction, Section 2 presents the family of aperture filters. Section 3 recalls the design of aperture filters. Section 4 presents the methodology of resolution constraining an operator. Section 5 presents the methodology of multiresolution design of operators. Section 6 shows some examples of application. Finally, Section 7 gives some conclusions.

2. Windowed Operators

Digital signals (or images) can be formally defined and represented by functions from a non-empty set E, that is an Abelian group with respect to a binary operation $+$, to an interval L . Usually, E is a subset of the set Z of integers ($Z \times Z$ for images) and L is the interval $[0, l-1]$ with $l \in Z^+$. A binary image is an element of the collection of subsets of E, denoted $\mathcal{P}(E)$. It can also be represented as a function from E to $[0,1]$ via the indicator function [23]. The set of all possible functions from E to L will be denoted by L^E. A mapping Ψ from L^E to L'^E will be called an image *operator* or *filter*, where L' is the interval $[0, l'-1]$ with $l' \in Z^+$.

2.1. W-Operators

A finite subset W of E, will be called a *window* and the number of points in W will be denoted by $|W|$. A *configuration* is a function from W to L and the space of all possible configurations from W to L will be denoted by L^W. Configurations usually result from translating a window W by t, $t \in E$, and observing the values of a signal h, $h \in L^E$, within the translated window, W_t. If $W = \{w_1, w_2, \ldots, w_n\}$, $n = |W|$, and we associate the points of W to an n-tuple $(w_1, w_2, \ldots, w_n)$, then a configuration $h(W_t)$ is given by

$$h(W_t) = (h(t + w_1), h(t + w_2), \ldots, h(t + w_n)).$$

Digital signals can be modeled by digital random functions and, in this sense, $h(W_t)$ is a realization of a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, that is, $h(W_t) = \mathbf{x} = (x_1, x_2, \ldots, x_n)$, where $\mathbf{x}$ denotes a realization of $\mathbf{X}$. An important subclass of operators from L^E to L'^E is the class of *W-operators*. They are translation invariant (t.i.) and locally defined (l.d.) within W. If an image operator Ψ is a W-operator, then it can be characterized by a function $\psi : L^W \to L'$, called a *characteristic function*, by [24, 25]

$$\begin{aligned}\Psi(h)(t) &= \psi(h(t + w_1), h(t + w_2), \ldots, h(t + w_n)) \\ &= \psi(\mathbf{x}).\end{aligned}$$

2.2. Aperture Operators

An aperture configuration is a function from W to K ($K = [-k, k], k \in Z^+$), and the set of all possible aperture configurations on W is denoted by K^W. These configurations are usually the result of a spatial translation of a window W by t, $t \in E$, range translating W by z, $z \in Z$, and truncating the observed values to values inside K. In this case, a configuration can be written as

$$\begin{aligned}h^*_{-z}(W_t) = (&h^*_{-z}(t + w_1), h^*_{-z}(t + w_2), \ldots, \\ &h^*_{-z}(t + w_n)),\end{aligned}$$

where $h_{-z}(t) = h(t) - z$, $z = z(h(W_t))$ is a function from $h(W_t)$ to L, and h^*_{-z} is defined by

$$h^*_{-z}(t) = \begin{cases} h_{-z}(t) & : \quad -k \le h_{-z}(t) \le k \\ k & : \quad h_{-z}(t) > k \\ -k & : \quad h_{-z}(t) < -k \end{cases} \tag{1}$$

Besides being t.i. and l.d. within a window W, aperture operators are also locally defined by a range window $K = [-k, k]$. Let $\mathbf{X}^*$ be a truncated random variable $\mathbf{X}$ (following the truncation rule of Eq. 1),

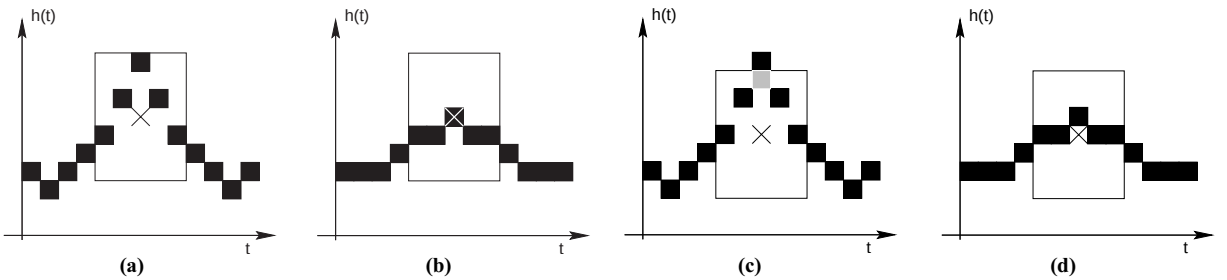


Figure 1. Aperture placement

that is, $\mathbf{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)$, and let $\mathbf{x}^*$ be a realization of $\mathbf{X}^*$. An aperture operator is defined by

$$\Psi(h)(t) = \psi(h^*_{-z}(t + w_1), h^*_{-z}(t + w_2), \ldots, h^*_{-z}(t + w_n)) + z = \psi(\mathbf{x}^*) + z,$$

where $z = z(h(W_t))$ is a function from $h(W_t)$ to L, and $\psi : K^W \to K$ is called the *characteristic function* of Ψ. It is important to notice that W-operators are aperture operators such that $k = l$ and $z = 0$.

Figure 1 shows two possible aperture placements. In part (a) and (b), corresponding to the ideal and observed signals, the aperture is placed at the observed values, that is, $z = h(t)$. Part (c) and (d) are analogous, with placement at the median in the observed signal, that is, $z =$ median $(h(W_t))$.

3. Design of Operators

Given two gray-level random images on E, h to be observed and g to be estimated, the basic design problem is to find an operator Ψ that minimizes an error measure between $\Psi(h)(t)$ and $g(t)$, where $t \in E$ [26]. The error measure to be minimized is usually the *Mean Absolute Error* (MAE) or the *Mean Square Error* (MSE). In the following sections, we present design techniques that minimize the MSE.

3.1. Design of Aperture operators

Assuming the gray-level random images h and g are jointly stationary [26], the MSE of Ψ is given by $E[(g(t) - \Psi(h)(t))^2]$. If Ψ is a W-operator, this is equivalent to $E[(Y - \psi(\mathbf{X}))^2]$, because Ψ is fully defined by its characteristic function ψ. We study operators by their equivalent characteristic functions. For ψ defined from L^W to L (the expression for a ψ defined from K^W to K is equivalent), its MSE is given by

$$\mathrm{MSE}(\Psi) = \sum_{\mathbf{x} \in L^W} \sum_{y=0}^{l-1} (y - \psi(\mathbf{x}))^2 P(y, \mathbf{x}) \qquad (2)$$

where $P(y, \mathbf{x})$ is the joint probability of $(y, \mathbf{x})$

The optimal MSE filter $\psi_{opt}(\mathbf{x})$ is given by $\lfloor E[Y|\mathbf{x}] + 0.5 \rfloor$, where $\lfloor \bullet + 0.5 \rfloor$ denotes the *floor* of $(\bullet + 0.5)$. This operation returns the nearest integer of $\bullet$. In practical applications, a design process is used to estimate ψ_{opt} as a function of samples obtained from observed-ideal image pairs. The result is an approximation of the optimal filter, that is, a *suboptimal* filter. We next analyze the MSE of a filter indirectly by its increment with respect to the MSE of the optimal filter, using $E[Y|\mathbf{x}]$ instead of $\lfloor E[Y|\mathbf{x}] + 0.5 \rfloor$ to simplify the calculus.

3.2. Error Analysis

When using a suboptimal filter instead of the optimal filter, there is an increase in error. The total increase in MSE error from using a given filter ψ in place of the optimal filter ψ_{opt} is [26]

$$\begin{aligned} \Delta(\psi, \psi_{opt}) &= \mathrm{MSE}(\psi) - \mathrm{MSE}(\psi_{opt}) \\ &= \sum_{\mathbf{x} \in L^W} (\psi(\mathbf{x}) - \psi_{opt}(\mathbf{x}))^2 P(\mathbf{x}) \\ &= E[(\psi(\mathbf{X}) - \psi_{opt}(\mathbf{X}))^2]. \end{aligned} \qquad (3)$$

This expression characterizes the MSE error increase of any filter as a function of the square distance between the filter and the optimal ideal filter. In

Appendix A, we analyze this error increase with respect to the optimal discrete valued filter.

If $\psi_{opt,N}$ is an estimate of ψ_{opt} based on N sample pairs $(\mathbf{X}^1, Y^1), \ldots, (\mathbf{X}^N, Y^N)$, then there is a design (estimation) cost $\Delta(\psi_{opt,N}, \psi_{opt})$, and the MSE of the designed filter can be given as a function of that cost and the MSE of the optimal filter, that is,

$$\mathrm{MSE}(\psi_{opt,N}) = \mathrm{MSE}(\psi_{opt}) + \Delta(\psi_{opt,N}, \psi_{opt}). \quad (4)$$

Since the estimated filter $\psi_{opt,N}$ depends on the training sample, it is random. Estimation error (*precision*) is defined by the expected cost $E[\Delta(\psi_{opt,N}, \psi_{opt})]$ and depends on the estimation procedure. The expected MSE of $\psi_{opt,N}$ is found by taking the expected value of Eq. 4, in which $\mathrm{MSE}(\psi_{opt})$ is constant, that is,

$$E[\Delta(\psi_{opt,N}, \psi_{opt})] = \sum_{\mathbf{x}\in L^W} \sum_{\lambda=0}^{l-1} (\lambda - \psi_{opt}(\mathbf{x}))^2 P[\psi_{opt,N}(\mathbf{x}) = \lambda] P(\mathbf{x}) \quad (5)$$

To estimate $E[Y|\mathbf{x}]$ and consequently $\psi_{opt,N}(\mathbf{x})$, we use the nearest integer to the sample mean $\overline{Y}_{\mathbf{x}}$, that is, $\psi_{opt,N}(\mathbf{x}) = [E[Y|\mathbf{x}] + 0.5]$. This is a consistent estimator. For a configuration unseen in the training sample, an alternative is to use the sample expectation of Y, that is, $\psi_{opt,N}(\mathbf{x}) = \lfloor \overline{Y} + 0.5 \rfloor$.

4. Resolution Constraint

This section presents the main results on resolution constraining an operator. We begin by introducing resolution constraint in the special case of down-sampling. We then consider the general definitions, first for spatial resolution, second for range resolution, and finally for the combination of both spatial and range resolution. The section is completed by analyzing the error increase of using resolution constraint.

4.1. Down-sampling Resolution Constraint

For the special case of down-sampling spatial resolution constraint, let W_0, W_1 be two windows such that $W_1 \subset W_0 \subset E$, $L = [0, \ldots, l-1]$ be the graylevel range, and $D_1 = L^{W_1}$ and $D_0 = L^{W_0}$ be the configuration spaces over W_1 and W_0 respectively. Let $\rho : D_0 \to D_1$ be a *sub-sampling*, that is, a mapping that assigns to each configuration $\mathbf{x} \in D_0$ a configuration $\mathbf{z} = \rho(\mathbf{x})$, which is a vector whose values are those of $\mathbf{x}$ over the sub-window W_1. Figure 2 shows a configuration $\mathbf{x} = (x_1, \ldots, x_{16})$ of W_0 that is sub-sampled to obtain a configuration $\mathbf{z} = (z_1 = x_1, z_2 = x_3, z_3 = x_9, z_4 = x_{11})$ over W_1, where $x_1, \ldots, x_{16} \in L$. Let L^{D_0} be the space of all functions from D_0 to L. The mapping ρ defines a constraint $\mathbf{Q} \subset L^{D_0}$ by: $\psi \in \mathbf{Q}$ if and only if (iff) $\rho(\mathbf{x}) = \rho(\mathbf{x}') \Rightarrow \psi(\mathbf{x}) = \psi(\mathbf{x}')$, for any $\mathbf{x}, \mathbf{x}' \in D_0$. The size of the configuration space is reduced for the constrained filter because $|D_1| = |L|^{|W_1|}$ and $|D_0| = |L|^{|W_0|} = |L|^{(|W_0|-|W_1|+|W_1|)} = |L|^{(|W_0|-|W_1|)}|D_1|$. Using the mapping ρ, an optimal constrained operator in L^{D_0} is estimated via the optimal operator in L^{D_1}.

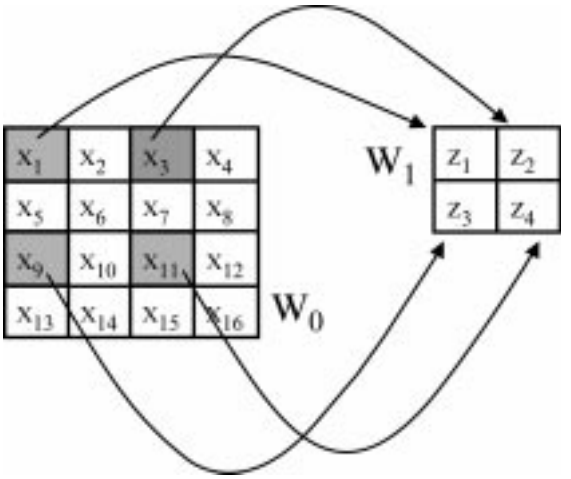


Figure 2. Downsampling.

For range resolution constraint, consider window $W \subset E$ and integer intervals $L_0 = L = [0, \ldots, l_0 - 1]$ and $L_1 = [0, \ldots, l_1 - 1]$, where l_0 is an even number and $l_1 = \frac{l_0}{2}$, be the gray-level ranges for the observation spaces $D_0 = L_0^W$ and $D_1 = L_1^W$. Let $\rho : D_0 \to D_1$ be a *quantization*, that is, a mapping that assigns to each configuration $\mathbf{x} \in D_0$ a configuration $\mathbf{z} = \rho(\mathbf{x}) = \lfloor \frac{\mathbf{x}}{2} \rfloor \in D_1$, where $\lfloor \bullet \rfloor$ denotes the floor of $\bullet$. In this case, a configuration $\mathbf{x} \in D_0$ with up to l_0 gray levels is quantized to half the gray levels to obtain the configuration $\mathbf{z} \in D_1$ with up to $\frac{l_0}{2}$ gray levels. The size of the configuration space is reduced for the constrained filter because $|D_1| = |L_1|^{|W|}$ and

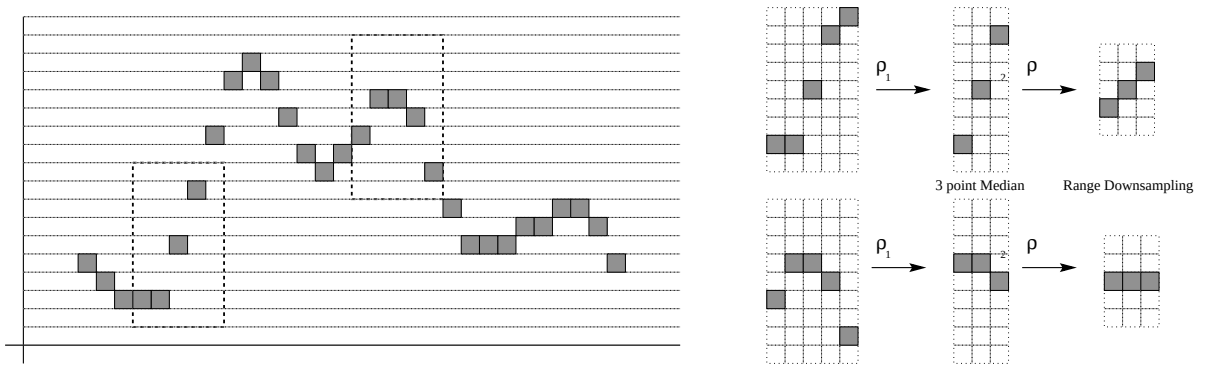


Figure 3. Combined spatial and range constraint.

$|D_0| = |L_0|^{|W|} = (2|L_1|)^{|W|} = 2^{|W|}|L_1|^{|W|} = 2^{|W|}|D_1|$. Analogous comments and analysis apply to other quantizations than $l_1 = \frac{l_0}{2}$.

Spatial and range constraints can be combined to obtain a new space that is a constraint of $D = L^W$. Consider windows $W_1 \subset W_0$ and let $L_0 = L = [0, \ldots, l_0 - 1]$ and $L_1 = [0, \ldots, l_1 - 1]$, with l_0 an even number and $l_1 = \frac{l_0}{2}$. Let $D_0 = L_0^{W_0}$, $D_{01} = L_0^{W_1}$, $D_1 = L_1^{W_1}$, and $\rho_1 : D_0 \to D_{01}$ be defined by spatial resolution constraint from W_0 to W_1. Let $\rho_2 : D_{01} \to D_1$ define a resolution constraint over W_1 by $\rho_2(\mathbf{x}) = \lfloor \frac{\mathbf{x}}{2} \rfloor$, for any $\mathbf{x} \in D_{01}$. The composition $\rho : D_0 \xrightarrow{\rho_1} D_{01} \xrightarrow{\rho_2} D_1$ is a mapping from D_0 to D_1 which takes a configuration $\mathbf{x} \in L_0^{W_0}$, subsamples and quantizes it to obtain a configuration $\mathbf{z} = \rho(\mathbf{x}) \in L_1^{W_1}$. The mapping ρ defines a resolution constraint on the space L^{D_0}. Figure 3 shows an example of combined spatial and range constraint. In this case, $L_0 = [-4, -3, -2, -1, 0, 1, 2, 3, 4]$, $L_1 = [-2, -1, 0, 1, 2]$, W_0 is a 5-point vector, W_1 is a 3-point vector, and

$$\rho_1(x_1, x_2, x_3, x_4, x_5) = (\text{median}(x_1, x_2, x_3), \text{median}(x_2, x_3, x_4), \text{median}(x_3, x_4, x_5))$$

$$\rho_2(z_1, z_2, z_3) = (\lfloor \frac{z_1}{2} \rfloor, \lfloor \frac{z_2}{2} \rfloor, \lfloor \frac{z_3}{2} \rfloor)$$

Down-sampling can be defined more generally [10]. Consider a subsampling where each pixel $\mu_i \in W_1$, $1 \le i \le |W_1|$, corresponds to a subwindow $W_{\mu_i} \subset W_0$, where the subwindows form a partition of W_0. A downsampling is defined via mappings $\zeta_i : L_0^{W_{\mu_i}} \to L_1$ that assign values to each pixel $\mu_i \in W_1$ as functions of the configuration $\mathbf{x} \in D_0$ restricted to W_{μ_i}. Formally, $\rho : D_0 \to D_1$ is defined by

$$\mathbf{z} = \rho(\mathbf{x}) = (\zeta_1(\mathbf{x}|_{W_{\mu_1}}), \ldots, \zeta_{|W_1|}(\mathbf{x}|_{W_{\mu_{|W_1|}}}))$$

This definition extends all the previous ones. For example, if $L_1 = L_0 = L$ and W_{μ_i} consists of the single pixel in W_0 that is in the same place as μ_i, then the mappings are equivalent to the previously defined spatial resolution constraint.

4.2. Resolution Constraint

To define resolution constraint in its general form, consider two configuration spaces D_0 and D_1 related by a surjective mapping $\rho : D_0 \to D_1$. ρ determines an equivalence relation on D_0 by $\mathbf{x} \sim \mathbf{x}'$ iff $\rho(\mathbf{x}) = \rho(\mathbf{x}')$. Therefore, for each $\mathbf{z} \in D_1$, there exists an equivalence class $C[\mathbf{z}]$, given by $C[\mathbf{z}] = \rho^{-1}(\mathbf{z})$. For gray-level operators over a fixed range L, optimal design is relative to the product space $L \times D_0$ with probability mass $P(y, \mathbf{x})$, $y \in L, \mathbf{x} \in D_0$. A probability mass is induced on D_1 by $P(\mathbf{z}) = P(\rho^{-1}(\mathbf{z}))$ and on $L \times D_1$ by

$$P(y, \mathbf{z}) = P(\{y\} \times \rho^{-1}(\mathbf{z})).$$

As in the binary case [10], an operator ϕ on D_1 induces an operator ψ_ϕ on D_0 by $\psi_\phi(\mathbf{x}) = \phi(\rho(\mathbf{x}))$ (Fig. 4). ψ_ϕ is *spatially resolution constrained*, in accordance with the function ρ: if $\rho(\mathbf{x}) = \rho(\mathbf{x}')$ then $\psi_\phi(\mathbf{x}) = \psi_\phi(\mathbf{x}')$. Conversely, if ψ is any operator on D_0 satisfying the resolution constraint, then it induces an operator on D_1 by $\phi_\psi(\mathbf{z}) = \psi(\mathbf{x})$, where $\mathbf{x}$ is any vector in $C[\mathbf{z}]$. Let L^{D_0} and L^{D_1} be the classes of gray-level operators on D_0 and D_1, respectively. The mapping $\phi \to \psi_\phi$ defines an injection $L^{D_1} \to L^{D_0}$. If

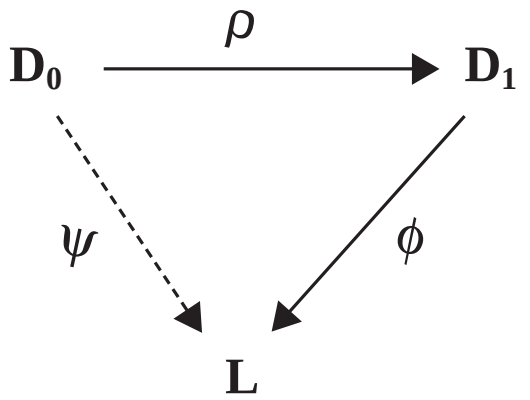


Figure 4. Commutative diagram.

Q is the subset of L^{D_0} composed of operators satisfying the resolution constraint, then the mapping $\psi \to \phi_\psi$ defines a bijection $\mathbf{Q} \to L^{D_1}$ whose inverse is given by the mapping $\phi \to \psi_\phi$. This bijective relation allow us to identify operators on D_1 with resolution constrained operators on D_0.

A key point here is whether an optimal operator on D_1 can be associated to an optimal resolution constrained operator on D_0. If this is the case, then consistent estimators for the optimal operator in L^{D_1} will induce consistent estimators for the optimal operator in **Q**. The following theorem shows this is always true for any loss function.

Theorem 4.1. *Error Preservation - Let* $\ell : L \times L \to \mathcal{R}^+$*, be a loss function used to define the risk functions* R*:* $L^{D_0} \to \mathcal{R}^+$*,* $R(\psi) = E[\ell(Y, \psi(\mathbf{X}))]$ *and* $R : L^{D_1} \to \mathcal{R}^+$*,* $R(\phi) = E[\ell(Y, \phi(\mathbf{Z}))]$*. Then* $R(\psi) = R(\phi_\psi)$*,* $\forall \psi \in \mathbf{Q}$*.*

Proof

$$\begin{aligned} R(\phi_\psi) &= \sum_{\mathbf{z}\in D_1}\sum_{y\in L} \ell[y, \phi_\psi(\mathbf{z})]P(y, \mathbf{z}) \\ &= \sum_{\mathbf{z}\in D_1}\sum_{y\in L} \ell[y, \phi_\psi(\mathbf{z})] \sum_{\mathbf{x}\in C[\mathbf{z}]} P(y, \mathbf{x}) \\ &= \sum_{\mathbf{z}\in D_1}\sum_{\mathbf{x}\in C[\mathbf{z}]}\sum_{y\in L} \ell[y, \phi_\psi(\mathbf{z})]P(y, \mathbf{x}) \\ &= \sum_{\mathbf{z}\in D_1}\sum_{\mathbf{x}\in C[\mathbf{z}]}\sum_{y\in L} \ell[y, \psi(\mathbf{x})]P(y, \mathbf{x}) \\ &= \sum_{\mathbf{x}\in D_0}\sum_{y\in L} \ell[y, \psi(\mathbf{x})]P(y, \mathbf{x}) \\ &= R(\psi) \end{aligned} \tag{6}$$

An immediate consequence of the Error Preservation Theorem (**EP**) is that, under spatial resolution constraint, MSE and MAE (*mean absolute error*) are preserved by the mapping $\psi \to \phi_\psi$ and, therefore, the optimal filter on D_1 induces the optimal resolution constrained filter on D_0.

4.3. Non-Preservation of Error in Range Resolution Constraint

The error-preservation theorem has been demonstrated under the assumption of resolution constraint in both the spatial domain and range, in particular with $D_0 = L_0^{W_0}$ and $D_1 = L_1^{W_1}$, and the condition that both ψ and the induced mapping ϕ_ψ possess the same range L, that is, $\psi \in L^{D_0}$ and $\phi_\psi \in L^{D_1}$. A natural situation is when ψ is being used to estimate the value Y of an ideal image based on the random vector **X** in an observation window, with $L = L_0$. Owing to error preservation, the optimal filter in **Q** is found by determining the optimal filter in L^{D_1}. Now consider an alternate situation in which resolution constraint is not only applied to the observation **X** but to the ideal value Y. Here, $\psi : D_0 \to L_0$, there exists a surjection $s : L_0 \to L_1$, and $\phi_\psi : D_1 \to L_1$. $\psi(\mathbf{X})$ serves as an estimator of Y, whereas $\phi_\psi(\mathbf{Z})$ serves as an estimator of $s(Y)$. The question now is whether, if ϕ is the optimal estimator for $s(Y)$ in L^{D_1}, does this guarantee that ψ_ϕ is the optimal estimator for Y in **Q**? As we will see, it does not.

To see the problem, consider the simplified situation in which there is no spatial constraint, so that there is a single window W, $D_0 = L_0^W$ and $D_1 = L_1^W$. Consider a surjection $s : L_0 \to L_1$ and a corresponding injection $r : L_1 \to L_0$, in which case, the composition $s \circ r$ is the identity on L_1. The pair (r, s) defines a constraint class $\mathbf{Q} \in L_0^{D_0}$ in accordance with the constraint mapping $\rho(\mathbf{x}) = (s(x_1), s(x_2), \ldots, s(x_{|W|}))$. $\psi \in \mathbf{Q}$ if and only if $\rho(\mathbf{x}) = \rho(\mathbf{x}')$ implies $\psi(\mathbf{x}) = \psi(\mathbf{x}')$, and $\psi(\mathbf{x}) \in r(L_1)$ for any $\mathbf{x} \in D_0$. An operator $\phi \in L_1^{D_1}$ induces a resolution-constrained operator $\psi_\phi \in L_0^{D_0}$ by $\psi_\phi(\mathbf{x}) = r(\phi(\rho(\mathbf{x})))$ [Fig. 5].

On the other hand, if ψ is a resolution constrained operator in $L_0^{D_0}$, then it induces an operator $\phi_\psi \in L_1^{D_1}$ by $\phi_\psi(\mathbf{z}) = s(\psi(\mathbf{x}))$ for any $\mathbf{x} \in D_0$ such that $\rho(\mathbf{x}) = \mathbf{z}$. The mapping $\phi \to \psi_\phi$ is a bijection from $L_1^{D_1}$ onto **Q** having inverse $\psi \to \phi_\psi$.

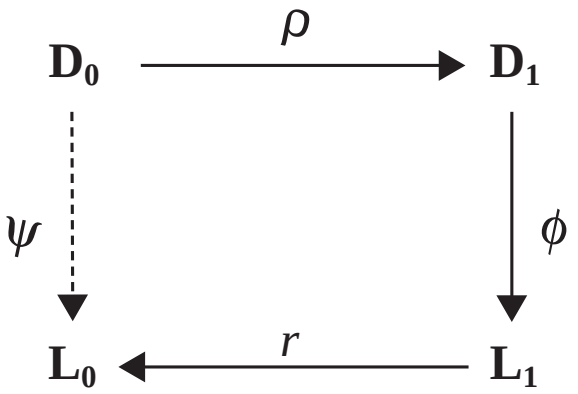


Figure 5. Commutative diagram.

Now consider the special case in which $s(j) = \lfloor \frac{j}{2} \rfloor$ and $r(i) = 2i$. The MSE for ψ as an estimator of Y and ϕ_ψ as an estimator of $s(Y) = \lfloor \frac{Y}{2} \rfloor$ are related by

$$\mathrm{MSE}(\phi_\psi) = \frac{1}{4}\mathrm{MSE}(\psi) + \frac{1}{4}\sum_{\mathbf{x}\in D_0}\sum_{i=0}^{l_1-1} 2[(\psi(\mathbf{x}) - 2i) - \frac{1}{2}]$$
$$P(2i+1, \mathbf{x}). \tag{7}$$

If $P(2i+1, \mathbf{x}) = 0$ for any $i \in L_1$, meaning the probability mass is concentrated in $r(L_1)$, then $\mathrm{MSE}(\phi_\psi) = \frac{1}{4}\mathrm{MSE}(\phi)$ and the optimal operator in $L_1^{D_1}$ induces the optimal operator in $\mathbf{Q}$; otherwise, it need not.

To exemplify this, let $l_0 = 4$ and $l_1 = 2$, then $L_0 = [0, 1, 2, 3]$ and $L_1 = [0, 1]$. Let $W = (u, v)$ be a two-point window, $\rho(\mathbf{x}) = (\lfloor \frac{x_1}{2} \rfloor, \lfloor \frac{x_2}{2} \rfloor)$. The space D_1 has $2^2 = 4$ different configurations: {00,01,10,11}. The space D_0 has $2^4 = 16$ different configurations: {00,01,02,03,10,11,12,13,20,21,22,23,30,31,32,33}. The equivalence classes are $C[00] = \{00, 01, 10, 11\}$, $C[01] = \{02, 03, 12, 13\}$, $C[10] = \{20, 21, 20, 31\}$ and $C[11] = \{22, 23, 32, 33\}$. There are $2^{2^2} = 16$ operators in the space $L_1^{D_1}$. These 16 operators induce (or are induced by) the 16 operators in $\mathbf{Q} \subset L_0^{D_0}$. Suppose now that the distribution P over $L_0 \times D_0$ is given by Table 6 (a), and that the induced distribution over $L_1 \times D_1$ is given by Table 6 (b). From these tables, it is straightforward to compute the MSE for each of the 16 operators in $L_1^{D_1}$ and for the 16 induced operators in $\mathbf{Q} \subset L_0^{D_0}$.

Table 6 (c) shows the MSE errors for the 16 constrained operators in $\mathbf{Q}$ (labeled MSE(ψ)) and for each of the 16 operators in $L_1^{D_1}$ (labeled MSE(ϕ_ψ)).

X	Y=0	Y=1	Y=2	Y = 3
00	0.01	0.02	0.01	0.01
01	0.01	0.00	0.02	0.01
02	0.02	0.02	0.03	0.01
03	0.01	0.01	0.01	0.02
10	0.02	0.01	0.02	0.02
11	0.01	0.01	0.02	0.02
12	0.01	0.03	0.03	0.01
13	0.02	0.00	0.02	0.03
20	0.03	0.02	0.01	0.02
21	0.01	0.01	0.02	0.01
22	0.00	0.02	0.01	0.01
23	0.01	0.02	0.03	0.02
30	0.02	0.01	0.01	0.01
31	0.03	0.02	0.00	0.02
32	0.01	0.01	0.03	0.01
33	0.01	0.02	0.02	0.02

(a)

X	Y=0	Y=1
00	0.09	0.13
01	0.12	0.16
10	0.15	0.10
11	0.10	0.15

(b)

Operator	$MSE(\psi)$	$MSE(\phi_\psi)$
1	3.6400	0.5400
2	2.9200	0.4900
3	3.3600	0.5900
4	2.6400	0.5400
5	2.9600	0.5000
6	2.2400	0.4500
7	2.6800	0.5500
8	1.9600	0.5000
9	3.0800	0.5000
10	2.3600	0.4500
11	2.8000	0.5500
12	2.0800	0.5000
13	2.4000	0.4600
14	1.6800	0.4100
15	2.1200	0.5100
16	1.4000	0.4600

(c)

Figure 6. (a) Distribution over $L_0 \times D_0$, (b) Distribution over $L_1 \times D_1$ (c) MSE for all the operators in the constraint and their induced operators.

The best MSE constrained operator in $\mathbf{Q}$ is ψ_{16}, which is defined by $\psi_{16}(\mathbf{x}) = 2$ for any $\mathbf{x} \in D_0$. It induces the operator $\phi_{\psi_{16}} \in L_1^{D_1}$ given by $\phi_{\psi_{16}}(\mathbf{z}) = 1$ for any

$\mathbf{z} \in D_1$. $\text{MSE}(\psi_{16}) = 1.4$ and $\text{MSE}(\phi_{\psi_{16}}) = 0.46$. However, the best MSE operator in $L_1^{D_1}$ is the operator 14. $\phi_{\psi_{14}}$ is defined by $\phi_{\psi_{14}}(00) = 1$, $\phi_{\psi_{14}}(01) = 1$, $\phi_{\psi_{14}}(10) = 0$ and $\phi_{\psi_{14}}(11) = 1$. It is induced by the operator ψ_{14} given by $\psi_{14}(\mathbf{x}) = 0$ if $\mathbf{x} \in C[10]$ and $\psi_{14}(\mathbf{x}) = 2$ for the remaining configurations $\mathbf{x} \in D_0$. $\text{MSE}(\phi_{14}) = 0.41$ and $\text{MSE}(\psi_{\phi_{14}}) = 1.68$.

The fact that the optimal operator in $\mathbf{Q}$, ψ_{16}, is different from the operator $\psi_{\phi_{14}}$, induced by the optimal operator in $L_1^{D_1}$, illustrates that we do not know the best operator in the constraint class $\mathbf{Q}$ through the best operator in the associated space $L_1^{D_1}$, even with the existence of a bijection between $\mathbf{Q}$ and $L_1^{D_1}$.

Nonetheless, many simulations have shown that the EP property either holds or is close to holding. Moreover, assuming it to be true has provided good practical results. Therefore we employ it in applications when there is a range change.

4.4. Error Analysis for Resolution Constraint

If $D_0 = L_0^{W_0}$ and $D_1 = L_1^{W_1}$ are the full and resolution-constrained spaces, respectively, and ψ_0 is the optimal (unconstrained) operator in L^{D_0}, then there are two possibilities concerning the optimal constrained operator ψ_1 in $\mathbf{Q}$, where $\psi_1 = \psi_{\phi_1}$ and ϕ_1 is optimal in L^{D_1}. The first possibility is that $\psi_0 \in \mathbf{Q}$, in which case $\psi_1 = \psi_0$ and there is no additional error owing to constraint. Of more interest is the second case, in which $\psi_0 \notin \mathbf{Q}$ and $\psi_1 \neq \psi_0$. Then there is a cost $\Delta(\psi_1, \psi_0)$ of constraint. The purpose of constraint is to lower the design cost and the worthiness of the constraint depends on the relation between the cost of constraint and the design costs for the constrained and unconstrained operators.

To examine this relation, suppose there are N sample pairs used to train the operator, and let $\psi_{0,N}$ and $\psi_{1,N}$ denote the operators that result from estimating the optimal unconstrained and constrained filters, respectively, from the sample data. According to Eq. 4, $E[\text{MSE}(\psi_{1,\text{N}})] \leq \text{E}[\text{MSE}(\psi_{0,\text{N}})]$ if and only if

$$E[\Delta(\psi_{0,N}, \psi_0) - \Delta(\psi_{1,N}, \psi_1)] \geq \Delta(\psi_1, \psi_0). \quad (8)$$

in which case the constraint is beneficial. The expectation of Eq. 8 depends on the way the operator is estimated.

As shown in Appendix C, Eq. 19, the constraint cost has the following expression:

$$\Delta(\psi_1, \psi_0) = E[\text{Var}[\text{E}[Y|\mathbf{X}] \ |\mathbf{Z}]] \quad (9)$$

For simulations, we have used $n_0 = 5$, $n_1 = 3$ and $L_0 = L_1 = L = \{0, 1\}$; and 2 distributions Π_1 and Π_2. The conditional probabilities $P_1(Y = 1|\mathbf{x})$ for Π_1 are defined from the conditional probabilities $P_1(Y = 1|\mathbf{z})$ in Table 7 (a) by $P_1(Y = 1|\mathbf{x}) = P_1(Y = 1|\mathbf{z})$ for $\mathbf{z}$ such that $\mathbf{x} \in C[\mathbf{z}]$. The equivalence class $C[\mathbf{z}]$ is defined by the mapping $\rho_2(x_1, x_2, x_3, x_4, x_5) = (x_2, x_3, x_4)$. For example, $C[(0,1,0)] = \{(0,0,1,0,0), (0,0,1,0,1), (1,0,1,0,0), (1,0,1,0,1)\}$ and therefore $P_1(Y = 1|(0, 0, 1, 0, 0)) = P_1(Y = 1|(0, 0, 1, 0, 1)) = P_1(Y = 1|(1, 0, 1, 0, 0)) = P_1(Y = 1|(1, 0, 1, 0, 1)) = 0.0421$. The conditional probabilities for distribution Π_2 are defined in the same way from the conditional probabilities of Table 7 (b) and the mapping $\rho_2(x_1, x_2, x_3, x_4, x_5) = (x_1, x_3, x_5)$. A third distribution Π_3 is a linear combination of Π_1 and Π_2: $P_3(y, \mathbf{x}) = 0.4P_1(y, \mathbf{x}) + 0.6P_2(y, \mathbf{x})$ with the necessary normalization to make Π_3 a probability distribution.

For each distribution, we have simulated samples of size $N = 10, 20, \ldots, 100$, and have trained the unconstrained and the constrained operators. Figures 8(a), 9(a) and 10(a) show the average MSE measured over 200 repetitions of the experiment for both operators, for Π_1, Π_2 and Π_3 respectively. In Figure 8(a), MSE (ψ_0) = MSE(ψ_1). For Π_1 the constrained operator performs better (lower MSE) than the unconstrained one. For Π_2, the distribution is well suited for estimation in D_0, and except for very small N, the constraint is not beneficial. For Π_3 the

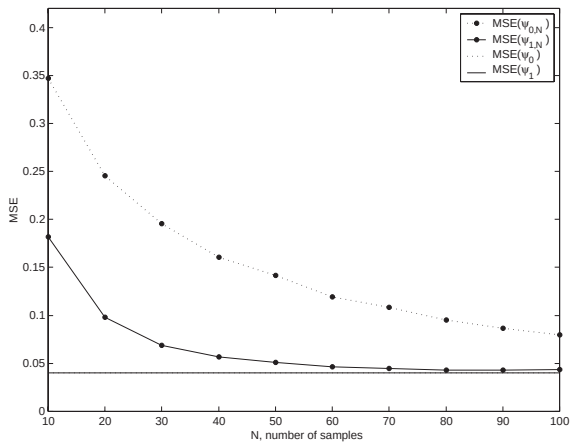
z	P(Y = 0)	P(Y = 1)
000	0.9579	0.0421
001	0.0421	0.9579
010	0.9579	0.0421
011	0.0421	0.9579
100	0.9579	0.0421
101	0.9579	0.0421
110	0.0421	0.9579
111	0.9579	0.0421

(a)

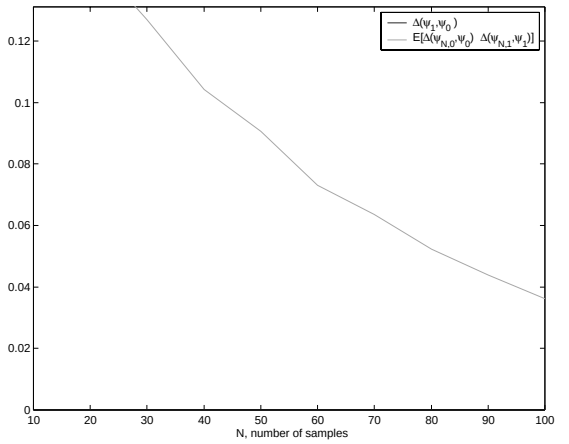
z	P(Y = 0)	P(Y = 1)
000	0.4198	0.5802
001	0.4963	0.5037
010	0.4643	0.5357
011	0.7088	0.2912
100	0.9579	0.0421
101	0.3544	0.6456
110	0.3472	0.6528
111	0.3280	0.6720

(b)

Figure 7. (a) Distribution Π_1 over $\{0, 1\} \times \text{D}_1$, (b) Distribution Π_2 over $\{0, 1\} \times \text{D}_1$.

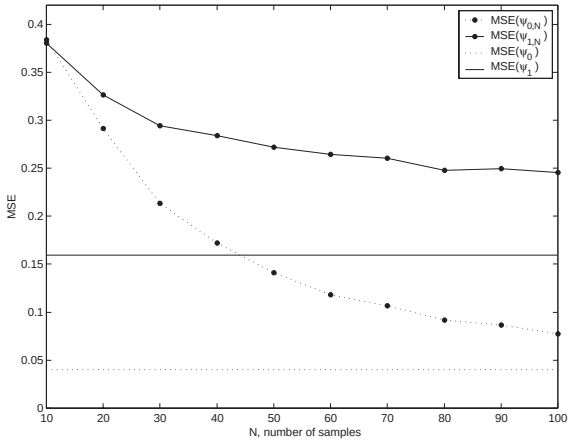


(a)

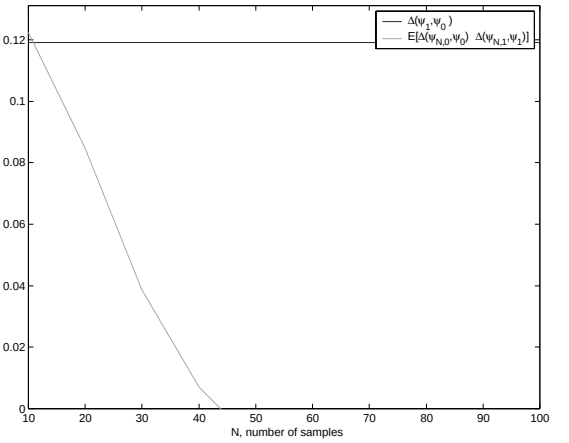


(b)

Figure 8. Π_1 a) MSE, b) Inequality



(a)



(b)

Figure 9. Π_2 a) MSE, b) Inequality

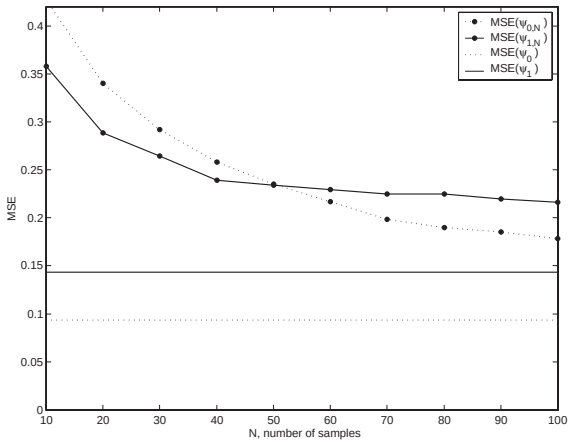
constrained distribution is better for less than 50 samples, but for a larger number of samples the unconstrained operator performs better.

Figures 8 (b), 9 (b) and 10 (b) show $\Delta(\psi_1, \psi_0)$ and the estimation of $E[\Delta(\psi_{0,N}, \psi_0) - \Delta(\psi_{1,N}, \psi_1)]$, for Π_1, Π_2 and Π_3, respectively. In Figure 8 (b), $\Delta(\psi_1, \psi_0) = 0$, because the optimal filter ψ_0 is a resolution constrained filter, and thus $\psi_0 = \psi_1$.

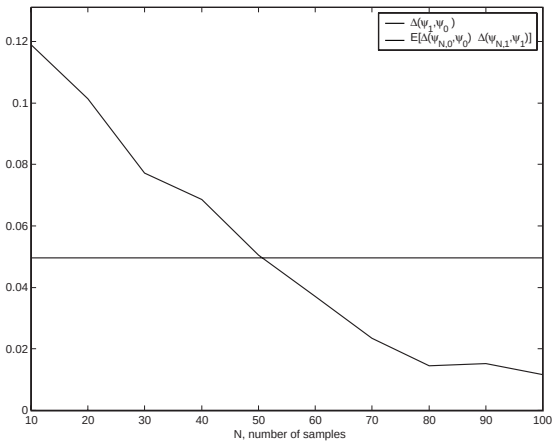
5. Multiresolution Design

In this section, we show how spatial resolution constraint can be iterated and some consequences of this approach. Consider windows $W_2 \subset W_1 \subset W_0$, where the configurations for W_1 and W_2 are obtained by downsampling from the configurations for W_0 and W_1, respectively. Let $D_0 = L^{W_0}$, $D_1 = L^{W_1}$ and $D_2 = L^{W_2}$ be the configuration spaces for the windows W_0, W_1 and W_2. Let $\rho_1 : D_0 \to D_1$ and $\rho_2 : D_1 \to D_2$ be the downsampling mappings defined by $\mathbf{z} = \rho_1(\mathbf{x})$ and $\mathbf{v} = \rho_2(\mathbf{z})$. The mappings ρ_1 and ρ_2 define a mapping $\rho_{12} : D_0 \to D_2$ by $\rho_{12} = \rho_2 \circ \rho_1$.

More generally, let D_0, D_1 and D_2 be configuration spaces related by the surjective mappings $\rho_1 : D_0 \to D_1$ and $\rho_2 : D_1 \to D_2$. ρ_1 induces an equivalence relation on D_0 by $\mathbf{x} \sim_1 \mathbf{x}'$ iff $\rho_1(\mathbf{x}) = \rho_1(\mathbf{x}')$. This equivalence relation defines



(a)



(b)

Figure 10. Π_1 a) MSE, b) Inequality

a partition P_1 of the space D_0, where $P_1 = \{C[\mathbf{z}] : \mathbf{z} \in D_1\}$ and $C[\mathbf{z}] = \rho_1^{-1}(\mathbf{z})$. In the same way, ρ_2 induces a partition $\{C[\mathbf{v}] : \mathbf{v} \in D_2\}$ on the space D_1. The mapping $\rho_{12} : D_0 \xrightarrow{\rho_1} D_1 \xrightarrow{\rho_2} D_2$ induces another partition P_2 on D_0 by $\mathbf{x} \sim_2 \mathbf{x}'$ iff $\rho_{12}(\mathbf{x}) = \rho_{12}(\mathbf{x}')$. This partition is coarser than P_1, in the sense that $\rho_{12}^{-1}(\mathbf{v}) = \cup\{C[\mathbf{z}] : \mathbf{z} \in C[\mathbf{v}]\}$. Comparing constraint cost and estimation errors at different resolutions simply involves subtracting equations of the form given in Eq. 8. Treating the analysis in a general setting, multiresolution analysis for filter design results from recursive partitioning the configuration spaces D_0, D_1, D_2, ..., relative to recursively applied resolution mappings ρ_1, ρ_2, ..., where $\rho_k : D_{k-1} \to D_k$. For configuration space D_k, there corresponds a resolution constrained optimal filter ψ_k. The best filtering resolution for a sample size N is the one for which $\mathrm{MSE}(\psi_{k,N})$ is minimal.

5.1. Error Analysis for Multiresolution

According to Eq. 8, the multiresolution constraint is beneficial for $\rho_k \circ \rho_{k-1} \circ \ldots \circ \rho_1$ if

$$E[\Delta(\psi_{0,N}, \psi_0) - \Delta(\psi_{k,N}, \psi_0)] \geq \Delta(\psi_k, \psi_0) \quad (10)$$

When the multiresolution analysis is done by recursive partitioning, each ρ_k is a resolution mapping, and the optimal filters ψ_{k-1} and ψ_k are related by

$$\Delta(\psi_k, \psi_{k-1}) = E[\mathrm{Var}[E[Y|\mathbf{Z}]|\mathbf{W}]] \quad (11)$$

where $\mathbf{W} \in D_k$ and $\mathbf{Z} \in C[\mathbf{w}] = \{\mathbf{z} \in D_{k-1} : \rho_k(\mathbf{z}) = \mathbf{w}\} \subset D_{k-1}$ for each $\mathbf{w} \in D_k$. $\rho = \rho_k \circ \rho_{k-1} \circ \ldots \circ \rho_1$ is a resolution mapping between D_0 and D_k, and

$$\Delta(\psi_k, \psi_0) = E[\mathrm{Var}[E[Y|\mathbf{X}]|\mathbf{W}]] \quad (12)$$

where $\mathbf{W} \in D_k$ and $\mathbf{X} \in C[\mathbf{w}] = \{\mathbf{x} \in D_0 : \rho(\mathbf{x}) = \mathbf{w}\} \subset D_0$ for each $\mathbf{w} \in D_k$. Eq. 11 and 12 show two ways to compute the error increase for the optimal filter ψ_k on D_k; as the sum of the error increases for each mapping (Eq. 11), or directly as a function of the equivalence class defined over D_0 (Eq. 12).

5.2. Pyramidal Design of Optimal Resolution Constrained Operators

The resolution constrained design approach presented so far produces a function ψ_1 on D_1 and applies it on D_0 (where D_1 is a constrained space of D_0) by $\psi_0(\mathbf{x}) = \psi_1(\mathbf{z})$ if $\mathbf{z} = \rho(\mathbf{x})$, where $\rho : D_0 \to D_1$. That means that, even having data to estimate the filter relative to D_0, we only use data relative to D_1 via ρ.

An alternative approach is to use data from both spaces. If we have a good estimate of $p_{\mathbf{x}}$ (the probability distribution of $\mathbf{x}$) and $\psi_{0,N}(\mathbf{x}) \neq \psi_{1,N}(\mathbf{z})$, then it would be prudent to use $\psi_{0,N}(\mathbf{x})$ in place of $\psi_{1,N}(\mathbf{z})$. On the other hand, if we have a poor estimate, or no estimate, of $p_{\mathbf{x}}$, but a better estimate of $p_{\rho(\mathbf{x})}$, then it can be beneficial to use $\psi_{1,N}(\mathbf{z})$. An operator designed in this way is called a *multiresolution operator* because rather than applying only $\psi_{0,N}$ for all $\mathbf{x}$, or applying only $\psi_{1,N}$ for all $\mathbf{z}$, the precision of the probability estimates are considered and the function is chosen accordingly. In the simplest case, requiring only that $\mathbf{x}$ be observed at least once, the multiresolution estimator is given by,

$$\psi_{(0,1),N}(\mathbf{x}) = \begin{cases} \psi_{0,N}(\mathbf{x}), & \text{if} \quad N(\mathbf{x}) > 0 \\ \psi_{1,N}(\mathbf{x}), & \text{if} \quad N(\mathbf{x}) = 0 \end{cases} \tag{13}$$

This idea can be repeated for any number of constraints, applied iteratively by a sequence of resolution constraint operators $\rho_1, \rho_2, \ldots, \rho_m$, ultimately, until the size of W_i ($W_i \subset W_{i-1}$) is 1 and the size of L_i ($L_i \subset L_{i-1}$) is 2. In this case, the multiresolution estimator is given by,

$$\psi_{(0,\ldots,m),N}(\mathbf{x}) = \begin{cases} \psi_{0,N}(\mathbf{x}), & \text{if } N(\mathbf{x}) > 0 \\ \psi_{1,N}(\mathbf{x}), & \text{if } N(\mathbf{x}) = 0, \\ & \quad N(\rho_1(\mathbf{x})) > 0 \\ & \vdots \\ \psi_{m-1,N}(\mathbf{x}), & \text{if } N(\mathbf{x}) = 0, \ldots, \\ & \quad N(\rho_{m-2}(\mathbf{x})) = 0, \\ & \quad N(\rho_{m-1}(\mathbf{x})) > 0 \\ \psi_{m,N}(\mathbf{x}), & \text{if } N(\mathbf{x}) = 0, \ldots, \\ & \quad N(\rho_{m-1}(\mathbf{x})) = 0 \end{cases} \tag{14}$$

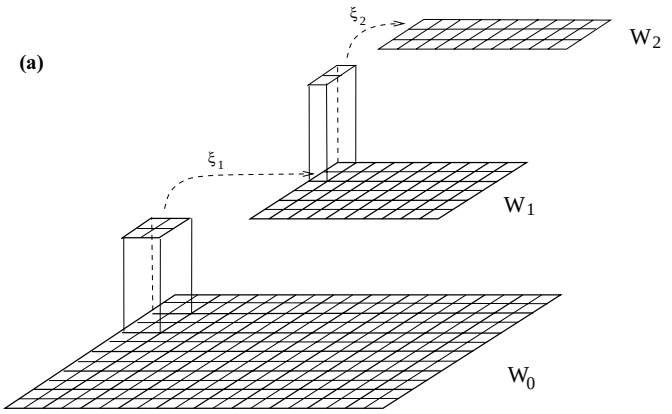
Figure 11a shows an example where ρ_0 maps each four pixels from window W_0 to one pixel in W_1 and ρ_1 maps each two pixels from window W_1 to one pixel in W_2. This is a typical case for binary or grayscale spatial multiresolution design of operators. A slight modification of this approach requires that $N(\mathbf{x})$ be sufficiently large (i.e., $\psi_{(0,1),N}(\mathbf{x}) = \psi_{0,N}(\mathbf{x})$, if $N(\mathbf{x}) > \alpha$, where $\alpha \in Z^+$).

Although the optimal filter at the higher resolution is better, the designed filter using multiresolution can outperform the standard designed filter owing to better probability estimates. In this sense, the pyramidal multiresolution design is a type of machine learning algorithm, where there is induction (generalization) [27] given by the successive constraints.

5.3. *Software Implementation for Pyramidal Multiresolution Design*

The implementation of a system to design and apply multiresolution operators is straightforward. It consists of three separate procedures: (*i*) *observation or estimation*, (*ii*) *decision and* (*iii*) *application*.

- The estimation procedure (*i*) consists of library functions to read pairs of images (observed, ideal), read the pyramid specification (the set of windows and range restrictions which compose the operator), slide the largest window (i.e., window W_0) on the set of training images and storing, for each configuration, the values observed in the ideal image for that configuration plus the number of times the configuration has been observed with that value. After observing and storing the configurations and observed values for the largest window, the same has to be done for all the windows in the multiresolution pyramid, but now it is enough to apply the constraint operators ρ_j to $\mathbf{x}$. The result of this procedure is a set of tables, one for each constrained operator ψ_j, where each line of each table contains one observed configuration and the values associated with their frequencies.
- The decision procedure (*ii*) has to compute only one value for each configuration (because an observed configuration can be linked to several ideal values) that minimizes a certain risk function $\mathcal{R}$; this has to be done for each table. The result of this procedure is a set of configuration tables (one for each constrained operator), where each line of each table contains one configuration, the respective decided value and the total of times the configuration has been observed.
- The application procedure (*iii*) uses functions to read an image, read the pyramid specification, read the result of the decision procedure, slide the window on the application image to get a configuration, compute the value to be assigned to the configuration and save the resulting image. This procedure is similar to the first one. The main difference is that, instead of storing configurations



1	2	1
2	3	2
1	2	1

(b)

Figure 11. (a) Pyramidal structure, (b) Convolution kernel.

and their respective associated values, the observed configuration will be searched in the table of configurations. If the configuration is found, the corresponding label found in the table is assigned to the pixel position in the resulting image. If the configuration is not found, the procedure applies each constraint operator ρ_j, $1 \leq j \leq m$, to the configuration, in this order, and repeats the searching in the corresponding table j. If the pyramid is well chosen, there will always be a table where the constrained configuration is found; otherwise, we can attribute a specific value to the output pixels.

6. Examples of Application

In this section, we show some examples comparing the multiresolution design to other known techniques. We have three possibilities for multiresolution design of aperture filters: (i) constrain the window, (ii) constrain the gray-scale and (iii) constrain both.

6.1. Deblurring 2D Images: Small Range

This application concerns deblurring of a random Boolean function model [28] whose primary function is pyramidal, with at most 16 gray levels. Blurring is accomplished by a (3×3) non flat convolution kernel (Fig. 11b). We have generated 20 images, each of size 256×256 points, and their respective blurrings. Figure 12 shows part of an original and blurred image. From this set, we have used 10 images for training and 10 for testing. Multiresolution aperture operators have been designed using the training set and then tested against the test set. Comparisons have

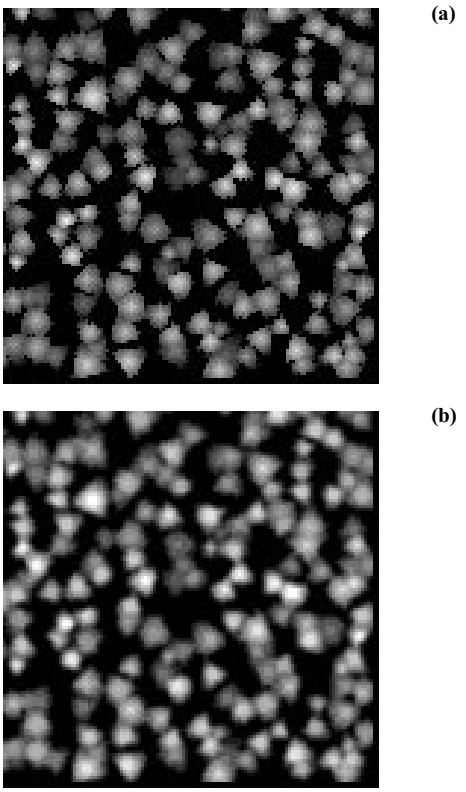


Figure 12. Random Boolean function: part of one image (a) and its blurring (b).

also been made with some previous experiments presented in Ref. [7] relative to some optimal linear filters, and some aperture operators without multiresolution that have been designed using the same training set and tested against the same test set. Figure 13 shows the MSE errors for some of the best designed operators. Each curve represents an experiment with a certain constraint. Each point of a curve represents the MSE of the operator (averaged over 10 test images) when using a certain amount of training examples to its design.

The curve labeled "blurred" is the error between the blurred and the ideal images. The curve labeled $Lin9 \times 9$ is the result of the optimal linear filter designed using a 9×9 window. Figures 14a and b show the pyramids used to design the W-operators labeled $MresW1$ and $MresW7$, respectively. Figures 15a and b show the pyramids used to design the W-operators labeled $MresW8$ and $MresW9$, respectively. The curve labeled $DT17 \times 5$ is the aperture operator designed using the window of 17 points inside a 5×5 window, denoted W_0 in the pyramid shown in Figure 15a. The curves labeled $MresA1$, $MresA7$, $MresA8$ and $MresA9$ are the MSE errors for multiresolution aperture operators designed using the pyramids shown in Figures 14a, 14b, 15a

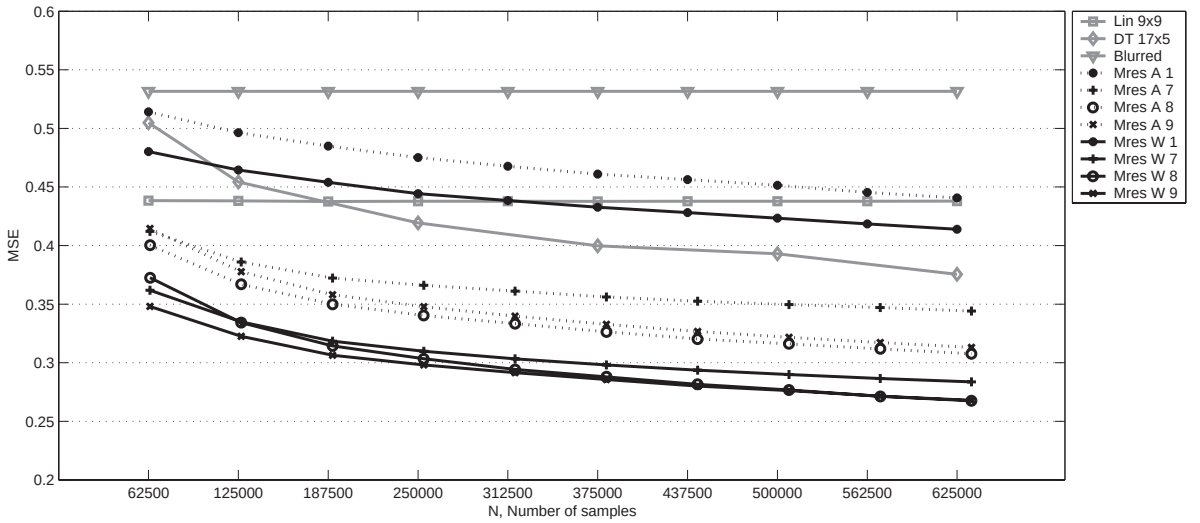


Figure 13. MSE error comparison: small range.

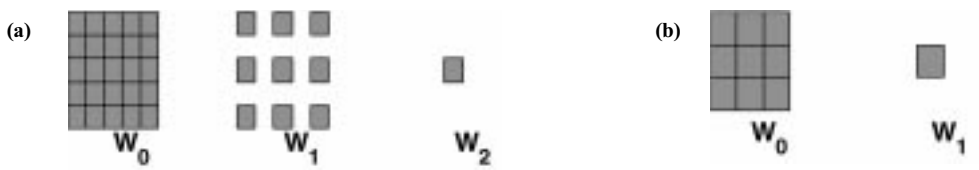


Figure 14. Pyramids for experiments "Mres W 1" and "Mres W 7".

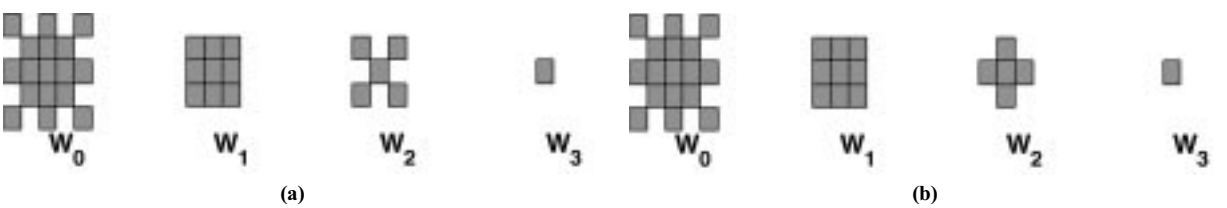
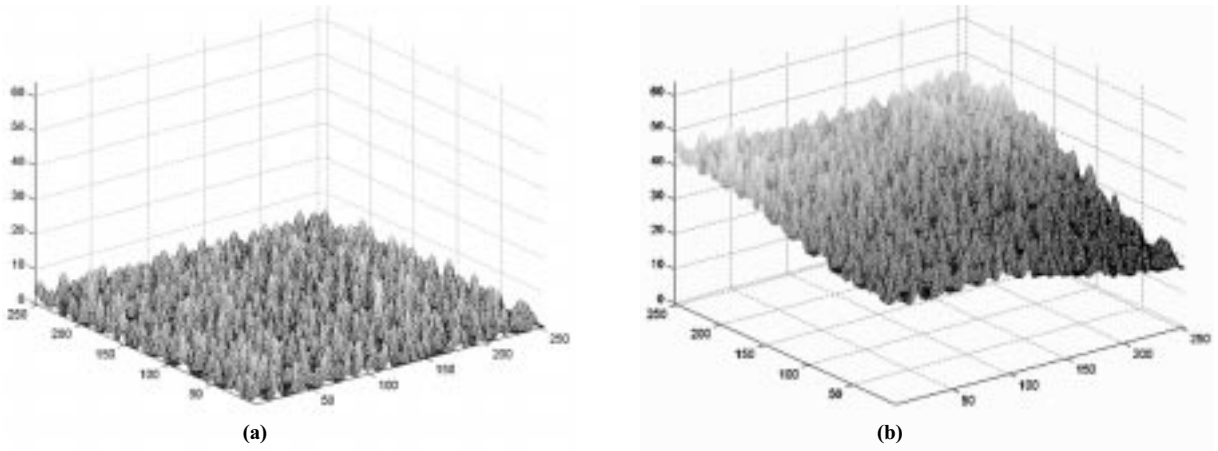


Figure 15. Pyramids for experiments "Mres W 8" and "Mres W 9".



(a) **(b)**

Figure 16. Surface of the random Boolean function: one image (a) combined with surface (b).

and 15b, respectively. Each aperture in the pyramid has a grayscale window range K, with $k = 5$.

Most of the non-linear filters perform better than the linear one, which stabilizes near 0.438, even though the linear filter is designed over a significantly larger window. Even at modest sample sizes, linearity is too strong of a constraint. The multiresolution filters *MresA*1 and *MresW*1 perform worse than the non-multiresolution aperture (17×5) filter, although for $N = 62500$, *MresW*1 outperforms the non-multiresolution aperture filter. The best results are for the filters *MresW*8 and *MresW*9; followed by *MresW*7, *MresA*9, *MresA*8 and *MresA*7. The Boolean function model used (short pyramids on a constant zero background) favors W-operators (aperture operators with $k = l$) over aperture operators with $k << l$, because the image range is relatively small and permits good estimation from the available data. The cost of the aperture constraint is not made up for by the savings in estimation error resulting from the aperture. But multiresolution design is certainly beneficial.

6.2. Deblurring 2D Images: Large Range

This application is similar to the previous one, but we have generated 20 different smooth functions to be summed to each of the 10 ideal images and to each of the 10 test images of the previous experiment. Each smooth function has been generated by choosing four random points at the corners of the image (uniform distribution between 0 and 50) and generating a smooth surface connecting those points. The final image is formed by placing small pyramids at different graylevel positions on a smooth image. The idea here is to model a pyramidal texture on a variable background. Figure 16 shows a 3D view of one original image (a) and the same image with the correspondent surface added (b) to it. The blurring has been generated using the same convolution kernel as previously and these new pairs (blurring and ideal) have been used to design new deblurring operators as before. Figure 17a shows part of one image resulting from adding one of the previous ideal images to one surface, and Figure 17b shows its blurring.

Figure 18 shows the MSE errors for some of the best designed operators in each class. It is constructed analogously to Figure 13. The curve labeled *blurred* is the error between the blurred and the ideal images. The curve labeled *Lin* 9×9 is the result of the optimal linear for a 9×9 points window. The curves labeled *MresA*6, *MresA*8, *MresA*9 and *MresW*8 show the MSE errors for multiresolution aperture and W-operators, respectively, designed using the pyramids shown in Figure 19a (*MresA*6), Figure 19b (*MresA*8 and *MresW*8) and Figure 20a (*MresA*9). Each aperture in the pyramid has the same grayscale range K, with $k = 5$. The curves labeled *DT*6 and *DT*15 show the MSE for the aperture operators designed using the square 3×3 window and the window of 17 points inside a 5×5 point window. The latter curve is

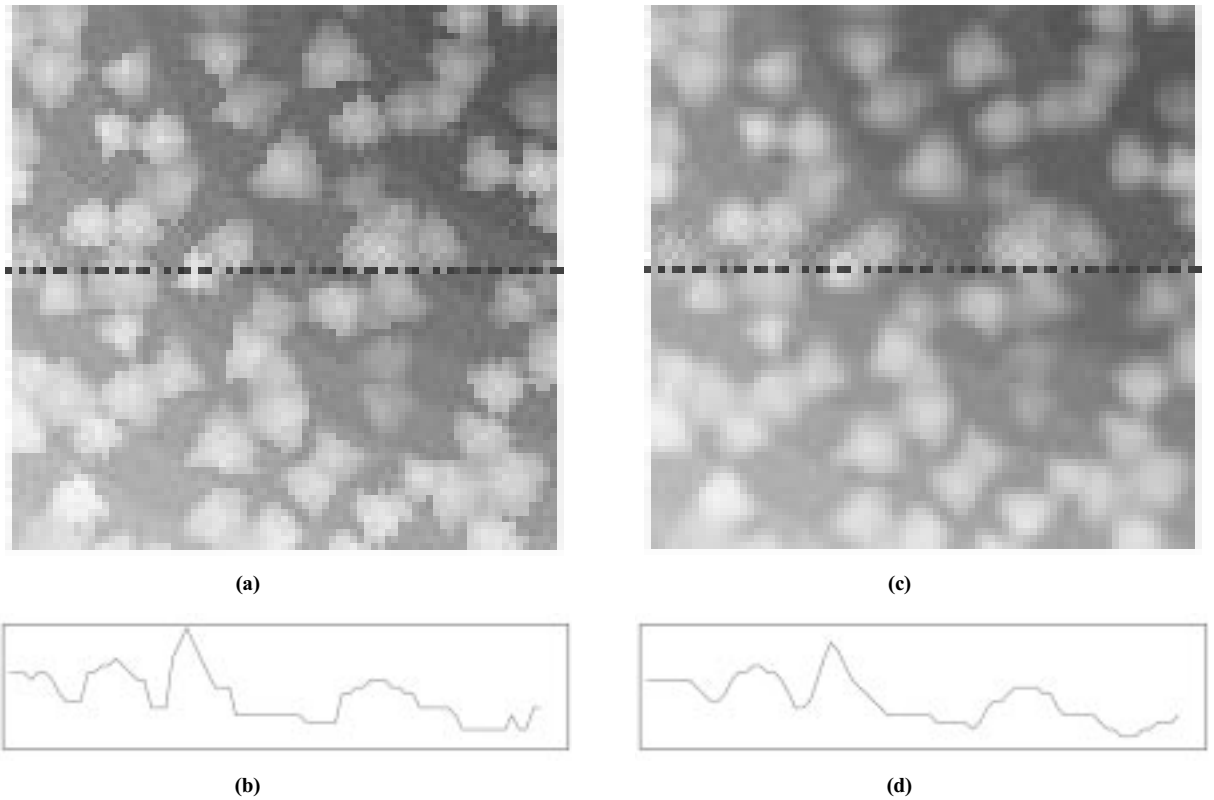


Figure 17. Random Boolean function: combined images (a) and a profile (b). Blurred image (c) and a profile (d).

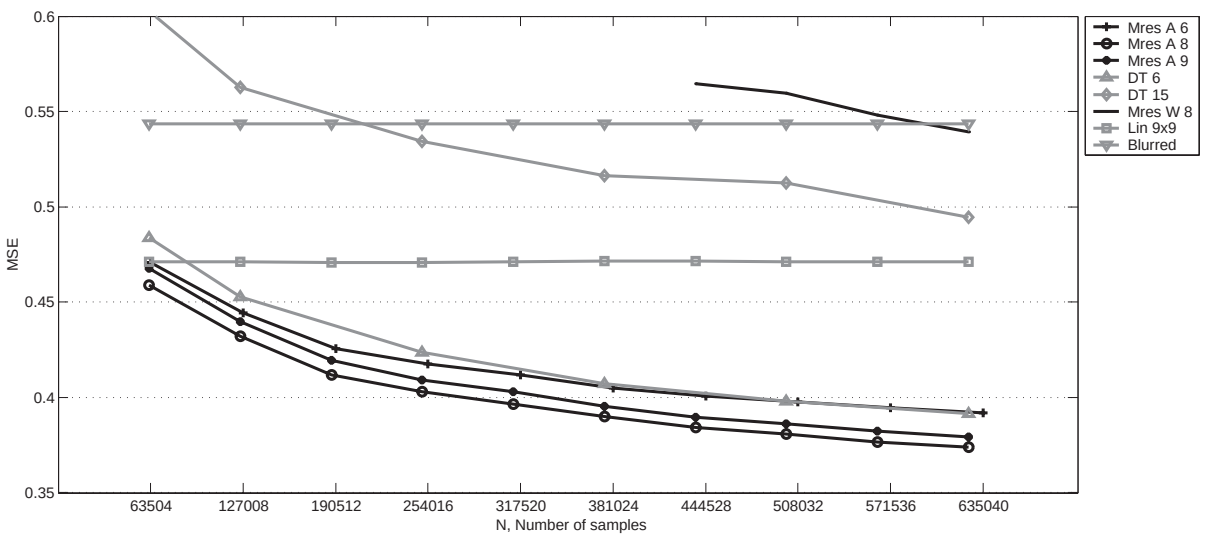
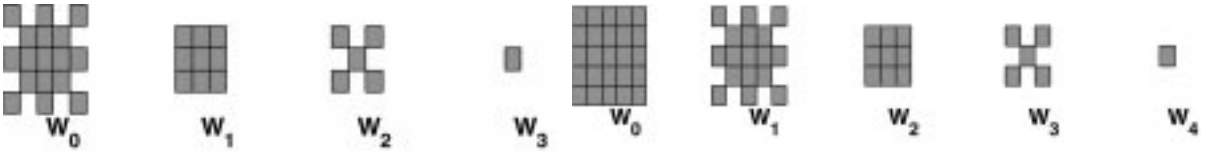


Figure 18. MSE error comparison: large range.

(a) (b)

Figure 19. Pyramids for experiments "6" and "8"



(a) (b)

Figure 20. Pyramids for experiments "9" and "11".

shown for comparison with the multiresolution aperture *MresA*8.

The best performing multiresolution W-operator is for *MresW*8, and its performance is very poor, with its MSE being worse than the MSE of the blurred image until $N = 635040$. The large gray range makes the cost of design exorbitant. The non-multiresolution aperture filter $DT15$ outperforms *MresW*8, but still requires too much sample data. Much better performance is achieved by the non-multiresolution aperture filter $DT6$. The multiresolution aperture filter *MresA*6 outperforms $DT6$ for smaller N, but $DT6$ catches up at $N = 381024$. The key point is that mutiresolution aperture filters *MresA*8 and *MresA*9 outperform all multiresolution W-filters and non-multiresolution aperture filters. The only difference between *MresA*8 and *MresA*9 is the window W_2 in their pyramids. In both cases window W_2 has 5 points, but experimentally the diagonal cross used for *MresA*8 works better than the horizontal-vertical cross used for *MresA*9 for the image model used here. A final note of interest is that the three multiresolution aperture filters considered in the figure all outperform the optimal linear filter for relatively small sample sizes. This demonstrates the ability of learned nonlinear filters to significantly outperform linear filters for deblurring large-range grayscale images.

The manner in which pyramidal multiresolution design provides improved performance for large windows is illustrated in Figure 21, which shows

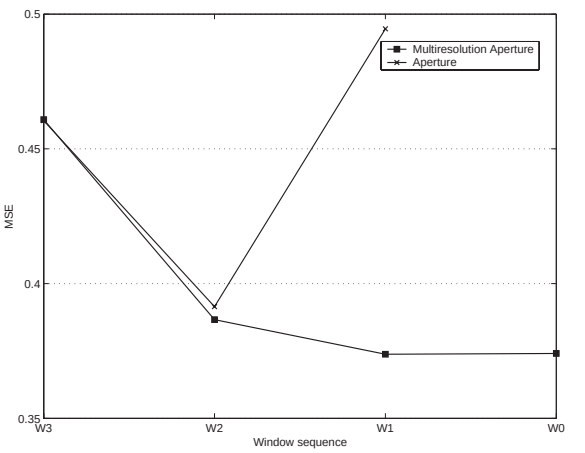


Figure 21. MSE error comparison: multiresolution × non-multiresolution.

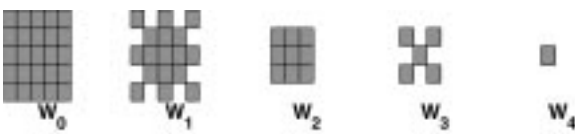


Figure 22. Pyramids: $\{W_3, W_4\}$, $\{W_2, W_3, W_4\}$,. . ., $\{W_0, W_1, W_2, W_3, W_4\}$.

MSE curves for multiresolution aperture operators and non-multiresolution aperture operators designed with the same number of samples. The multiresolution pyramid is composed of the windows $W_0, W_1, \ldots, W_4$ shown in Figure 22. For instance, the sequence

labeled W_2 is composed of the windows W_2, W_3 and W_4. The error curve shows that the best aperture operator for the given design samples is W_2. The error increases for larger or smaller windows. The error decreases for multiresolution aperture operators as more levels are added to the pyramid.

Figures 23 to 28 show a small region of a test image, its blurring, and the result of the best filter for each class (linear, multiresolution W-operator, non-multiresolution aperture, and multiresolution aperture). Figures 23 and 24 show a region of 500 points of the original image and the blurred image, respectively. The latter figure shows 161 points (marked by black edges) with different values. Most of them differ by 1 (129 points) or 2 (30 points). The estimated MSE for the region is 0.534. Figure 25 shows the result of the 9×9 linear filter. The MSE drops to 0.434, but the number of erroneous points does not decrease. The error decrease is due to the decrease of points that differ more than 1 (with a corresponding increase of the number of points that differ by 1). Figure 26 shows the result of the best multiresolution W-operator filter. In this case, the MSE rises to 0.552, but the number of erroneous points drop to 130. The number of points with difference 1 decreases to 102, while the number of points with difference greater than 1 increases. Figure 27 shows the result of the best non-multiresolution aperture filter (17 points window). The MSE and number of erroneous points drop to 0.3420 and 89, respectively. Especially visible is the improved restoration at grain edges. The best result is shown in Figure 28. It results from a multiresolution aperture filter starting with a 17-point window. The MSE and

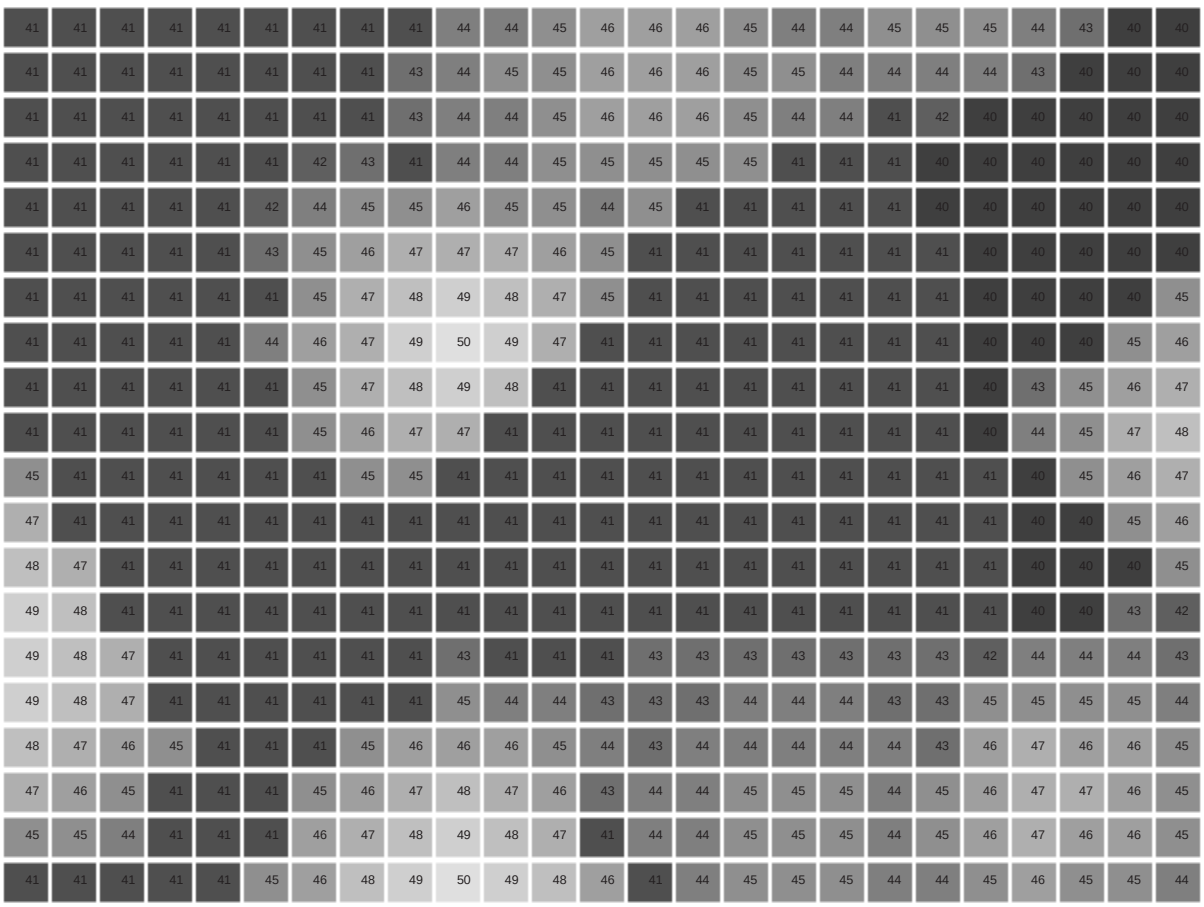


Figure 23. Section of the original Image.

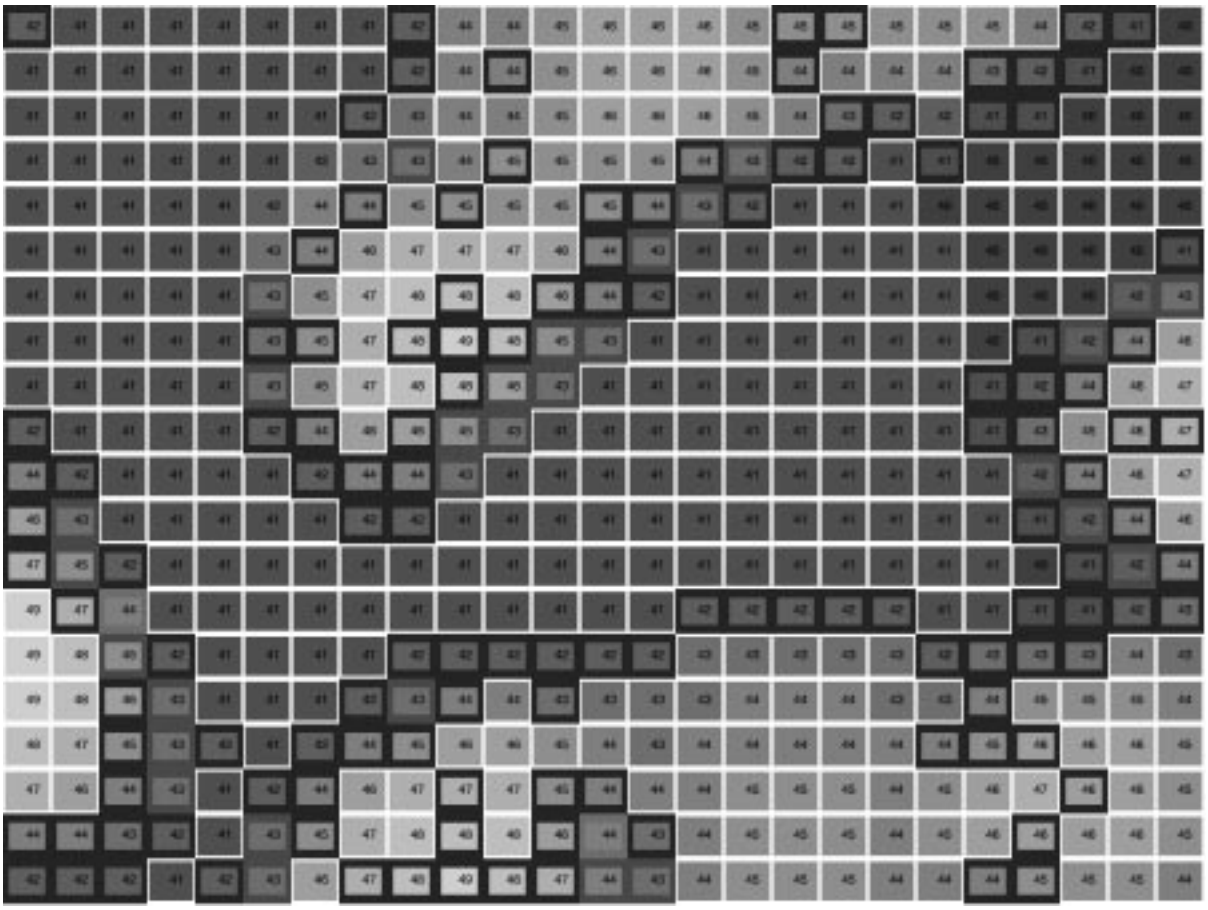


Figure 24. Blurred Image.

number of erroneous points drop to 0.222 and 73, respectively. Edge restoration is further improved over the non-multiresolution aperture filter.

7. Conclusion

By defining a window-based filter at the finest resolution for which it has been sufficiently observed, multiresolution design takes advantage of statistics at multiple levels and avoids the problem of generalizing the definition to vectors for which it has not been observed. This paper has extended the original binary methodology to gray-scale signals and images for both ordinary W-operators and aperture filters. It has empirically shown the advantage of multiresolution aperture design for large gray ranges. In particular, it has illustrated improved deblurring over optimal linear and non-multiresolution aperture filters.

Error preservation when the gray range is maintained has been proved, and empirical results have been presented to show that for practical purposes it is possible to design the filter in the resolution-constrained space even when the gray range is compressed. The errors of the multiresolution and pyramidal multiresolution approaches to aperture-filter design have been analyzed. The implementation of pyramid multiresolution design and application algorithms have been presented and discussed. Pyramid design remains heuristic in the sense that as of yet there is no automatic routine to design the constraints. It would be beneficial to be able to determine the best pyramid given knowledge concerning the filter class considered. Such an automatic approach must employ

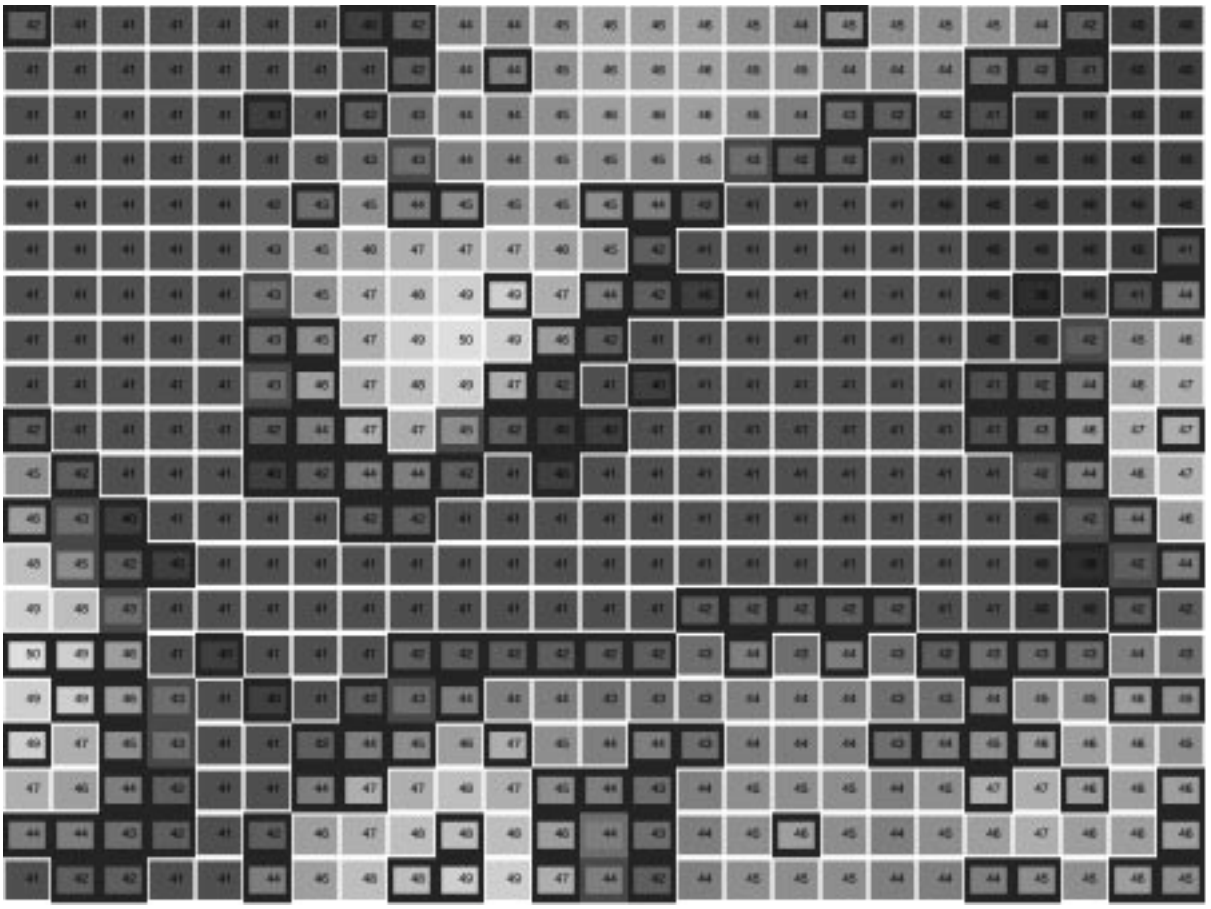


Figure 25. Result from best linear operator.

filter and image characteristics, since a strict combinatoric approach is not possible.

Appendix

A. *Quantized operator error analysis*

The value $\Delta(\psi, \psi_{opt})$ (Eq. 3) is useful to compare the error of two operators as error increases with respect to the optimal operator. In that analysis, $E[Y|\mathbf{x}]$ is used as the optimal operator for MSE error. If $\lfloor E[Y|\mathbf{x}] + 0.5 \rfloor$ is used, then the equations become more complex. Eq. 15 shows the expression for the error increase $\Delta(\psi, \psi_{opt})$ for the optimal quantized operator:

$$\begin{aligned}
\Delta(\psi, \psi_{opt}) &= \sum_{\mathbf{x}\in D}\sum_{y=0}^{l-1}(y - \psi_{opt}(\mathbf{x}))^2 P(y, \mathbf{x}) \\
&\quad - \sum_{\mathbf{x}\in D}\sum_{y=0}^{l-1}(y - \psi(\mathbf{x}))^2 P(y, \mathbf{x}) \\
&= \sum_{\mathbf{x}\in D}\sum_{y=0}^{l-1}[(\psi(\mathbf{x}) - \psi_{opt}(\mathbf{x}))^2 + 2(\psi(\mathbf{x}) \\
&\quad - \psi_{opt}(\mathbf{x}))(\psi_{opt}(\mathbf{x}) - y)]P(y, \mathbf{x}) \\
&= \sum_{\mathbf{x}\in D}[(\psi(\mathbf{x}) - \psi_{opt}(\mathbf{x}))^2 + 2(\psi(\mathbf{x}) \\
&\quad - \psi_{opt}(\mathbf{x}))(\psi_{opt}(\mathbf{x}) - E[Y|\mathbf{x}])]P(\mathbf{x}) \\
&= \sum_{\mathbf{x}\in D}[(\psi(\mathbf{x}) - \psi_{opt}(\mathbf{x}))^2 + 2\alpha_{opt}^{\mathbf{x}}(\psi(\mathbf{x}) \\
&\quad - \psi_{opt}(\mathbf{x}))]P(\mathbf{x}) \qquad (15)
\end{aligned}$$

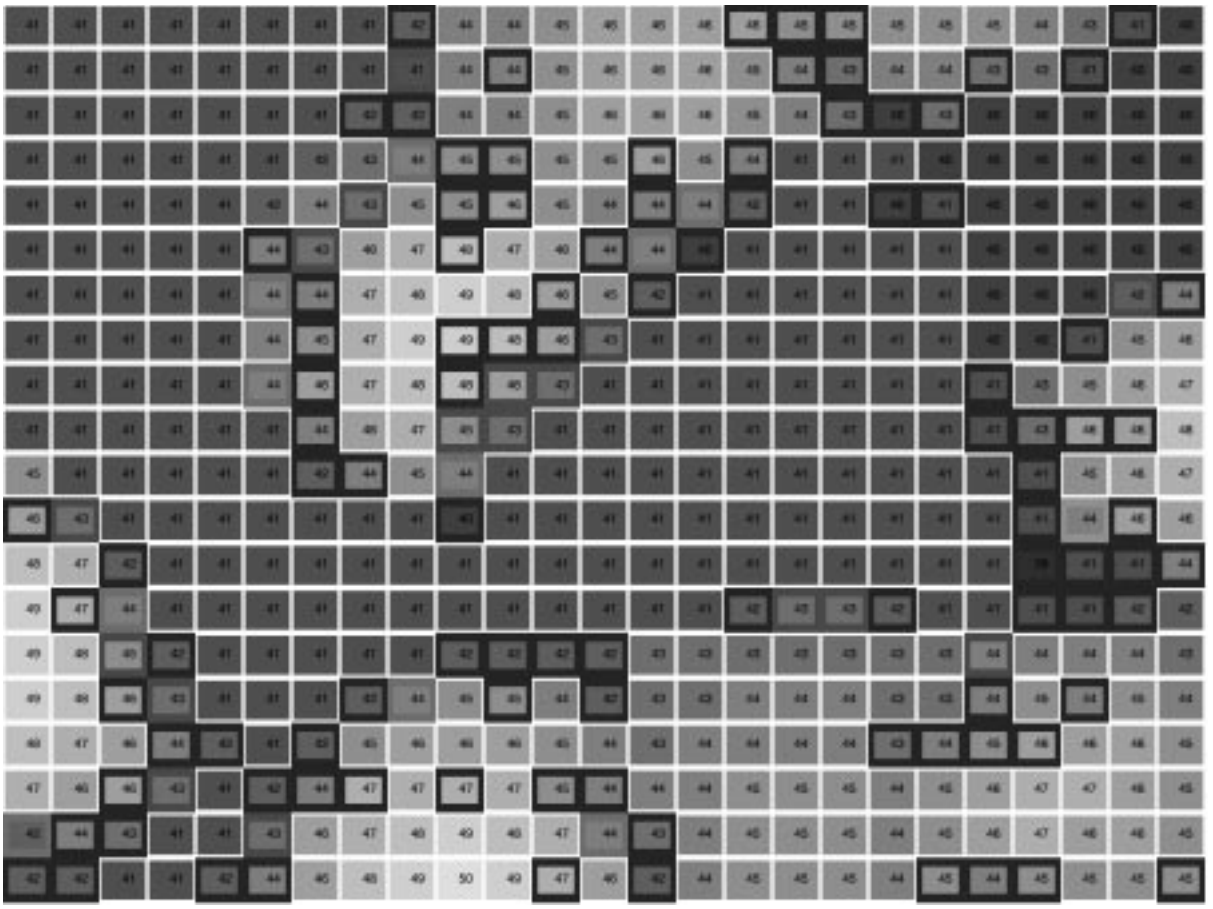


Figure 26. Result from best multiresolution W-operator.

where $\alpha_{opt}^{\mathbf{x}} = (\psi_{opt}(\mathbf{x}) - E[Y|\mathbf{x}])$ is the quantization error for ψ_{opt} at $\mathbf{x}$. Defining a polynomial function $A_{\mathbf{x}}(\lambda) = \lambda^2 + 2\alpha_{opt}^{\mathbf{x}}\lambda$ for each $\mathbf{x}$, the error increase can be written as,

$$\Delta(\psi, \psi_{opt}) = \sum_{\mathbf{x}\in D} A_{\mathbf{x}}(\psi(\mathbf{x}) - \psi_{opt}(\mathbf{x}))P(\mathbf{x})$$

B. *Quantizing the estimated value*

Proof of Eq. 7. Under the hypothesis of Eq. 7,

$$\begin{aligned}
\mathrm{MSE}(\phi_\psi) &= \sum_{\mathbf{z}\in D_1}\sum_{i=0}^{l_1-1}(i-\phi_\psi(\mathbf{z}))^2 P_1(i,\mathbf{z}) \\
&= \sum_{\mathbf{z}\in D_1}\sum_{i=0}^{l_1-1}(i-\phi_\psi(\mathbf{z}))^2 P_0(s^{-1}(i),\rho^{-1}(\mathbf{z})) \\
&= \sum_{\mathbf{z}\in D_1}\sum_{i=0}^{l_1-1}(i-\phi_\psi(\mathbf{z}))^2 \sum_{\mathbf{x}\in C[\mathbf{z}]}\sum_{j=0}^{1} P_0(2i+j,\mathbf{x}) \\
&= \sum_{\mathbf{z}\in D_1}\sum_{\mathbf{x}\in C[\mathbf{z}]}\sum_{i=0}^{l_1-1}\sum_{j=0}^{1}(i-\phi_\psi(\mathbf{z}))^2 P_0(2i+j,\mathbf{x}) \\
&= \sum_{\mathbf{x}\in D_0}\sum_{i=0}^{l_1-1}\sum_{j=0}^{1}(i-\frac{\psi(\mathbf{x})}{2})^2 P_0(2i+j,\mathbf{x}) \\
&= \frac{1}{4}\sum_{\mathbf{x}\in D_0}\sum_{i=0}^{l_1-1}\sum_{j=0}^{1}(2i-\psi(\mathbf{x}))P(2i+j,\mathbf{x}) \quad (16)
\end{aligned}$$

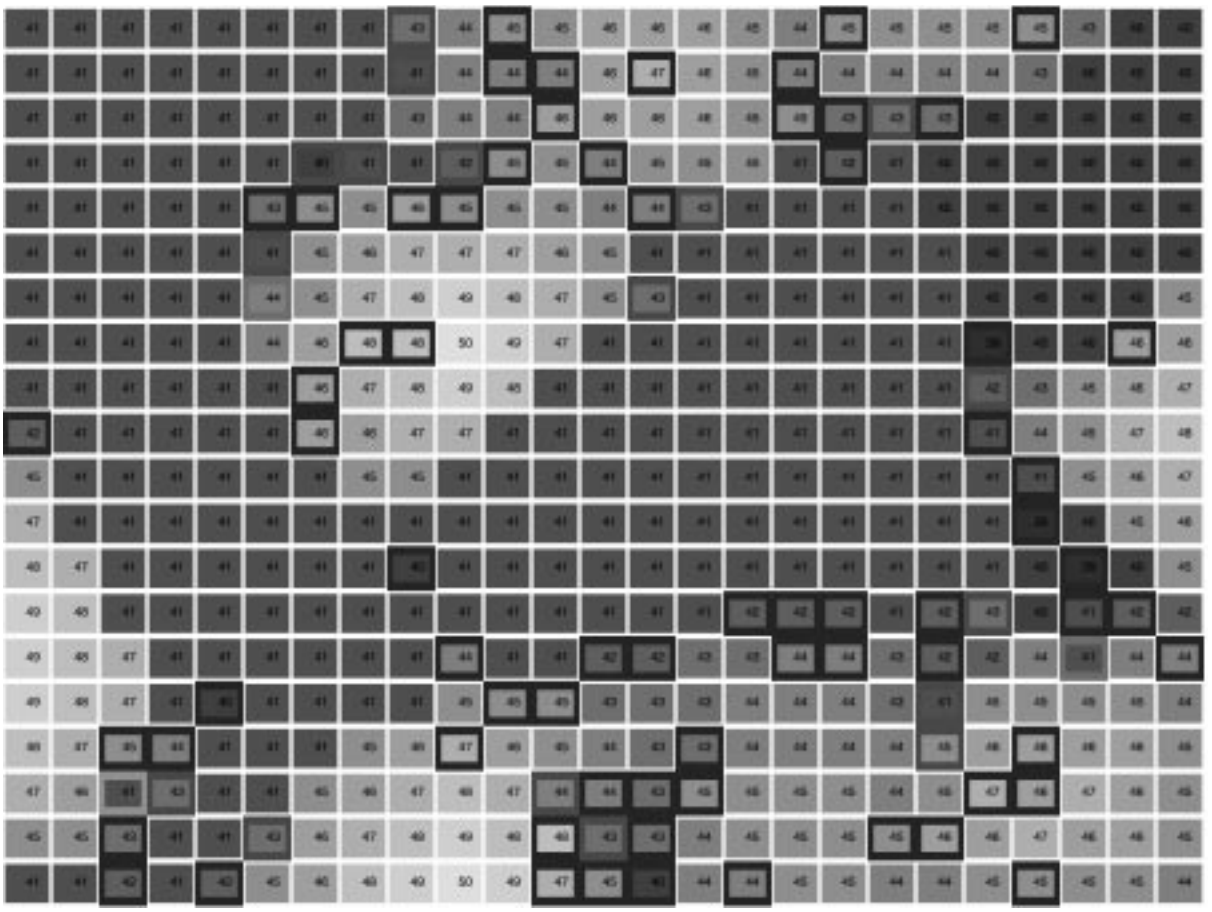


Figure 27. Result from best non-multiresolution aperture.

$$\begin{aligned}\mathrm{MSE}(\psi) &= \sum_{\mathbf{x}\in D_0}\sum_{t=0}^{l_0-1}(t-\psi(\mathbf{x}))^2P(t,\mathbf{x})\\ &= \sum_{\mathbf{x}\in D_0}\sum_{i=0}^{l_1-1}\sum_{j=0}^{1}(2i+j-\psi(\mathbf{x}))^2P(2i+j,\mathbf{x}) \quad (17)\end{aligned}$$

Hence, given that for $j=0$ the terms are identical,

$$\begin{aligned}\mathrm{MSE}(\phi_\psi)-\frac{1}{4}\mathrm{MSE}(\psi) &= \frac{1}{4}\sum_{\mathbf{x}\in D_0}\sum_{i=0}^{l_1-1}((2i-\psi(\mathbf{x}))^2\\ &\quad -(2i+1-\psi(\mathbf{x}))^2)P(2i+1,\mathbf{x})\\ &= \frac{1}{4}\sum_{\mathbf{x}\in D_0}\sum_{i=0}^{l_1-1}2[(\psi(\mathbf{x})-2i)\\ &\quad -\frac{1}{2}]P(2i+1,\mathbf{x}) \quad (18)\end{aligned}$$

Eq. 7 follows.

C. Constraint cost

$$\begin{aligned}\Delta(\psi_1,\psi_0) &= \sum_{\mathbf{x}\in D_0}(\psi_1(\mathbf{x})-E[Y|\mathbf{x}])^2P(\mathbf{x})\\ &= \sum_{\mathbf{z}\in D_1}\sum_{\mathbf{x}\in C[\mathbf{z}]}(\psi_1(\mathbf{z})-E[Y|\mathbf{x}])^2P(\mathbf{x}|\mathbf{z})P(\mathbf{z})\\ &= \sum_{\mathbf{z}\in D_1}E[(\psi_1(\mathbf{z})-E[Y|\mathbf{X}])^2|\mathbf{z}]P(\mathbf{z})\\ &= \sum_{\mathbf{z}\in D_1}(\mathrm{Var}[(\psi_1(z)-E[Y|\mathbf{X}])|\mathbf{z}]\\ &\quad +E^2[\psi_1(\mathbf{z})-E[Y|\mathbf{X}]|\mathbf{z}])P(\mathbf{z})\\ &= \sum_{\mathbf{z}\in D_1}(\mathrm{Var}[E[Y|\mathbf{X}]|\mathbf{z}]+(\psi_1(\mathbf{z})\end{aligned}$$

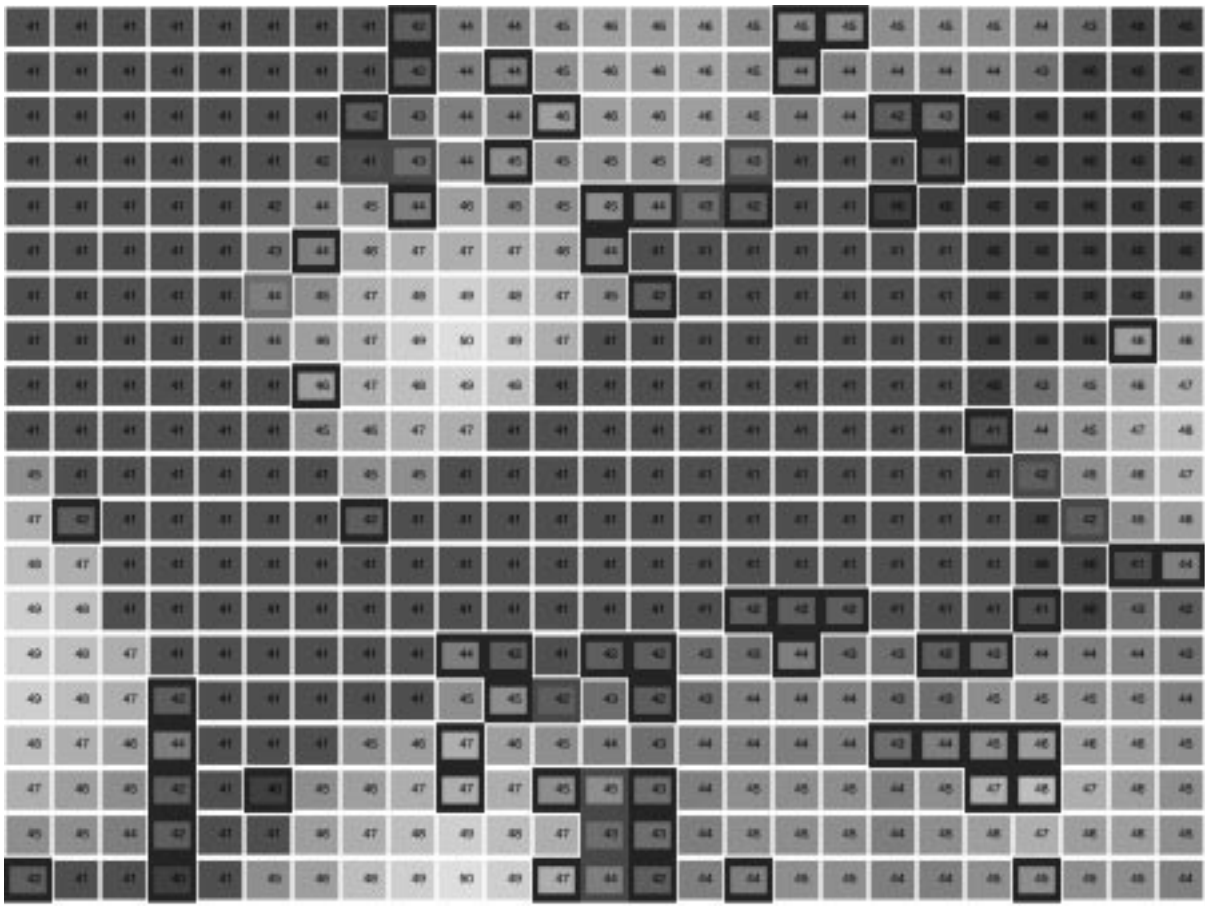


Figure 28. Result from best multiresolution aperture.

$$
\begin{aligned}
&= \sum_{\mathbf{z}\in D_1} (\mathrm{Var}[E[Y|\mathbf{X}]|\mathbf{z}] + (\psi_1(\mathbf{z}) \\
&\quad - E[E[Y|\mathbf{X}]|\mathbf{z}])^2)P(\mathbf{z}) \\
&= \sum_{\mathbf{z}\in D_1} (\mathrm{Var}[E[Y|\mathbf{X}]|\mathbf{z}] + (\psi_1(\mathbf{z}) \\
&\quad - E[Y|\mathbf{z}])^2)P(\mathbf{z}) \\
&= \sum_{\mathbf{z}\in D_1} \mathrm{Var}[E[Y|\mathbf{X}]|\mathbf{z}]P(\mathbf{z}) \\
&= E[\mathrm{Var}[E[Y|\mathbf{X}]|\mathbf{Z}]]. \qquad (19)
\end{aligned}
$$

Acknowledgments

The authors thank Dr. Artyom Grygoriam for the Boolean model images. Marcel Brun is supported by the FAPESP under grant 98/15586-9. Roberto Hirata Jr. was supported by CNPq under grant 145157/1998-9 during the development of this work.

References

1. E.R. Dougherty and J. Barrera. "Logical Image Operators," in E. R. Dougherty and J.T. Astola, editors, *Nonlinear Filters for Image Processing*, pp. 1–60. SPIE and IEEE Press, Bellingham, 1999.
2. E.R. Dougherty and J. Barrera, "Computatinal Gray-Scale Operators," in E.R. Dougherty and J.T. Astola, editors, *Nonlinear Filters for Image Processing*, pp. 61–98. SPIE and IEEE Press, Bellingham, 1999.
3. J. Barrera, E.R. Dougherty, and N.S. Tomita, "Automatic Programming of Binary Morphological Machines by Design of Statistically Optimal Operators in the Context of Computational Learning Theory," *Electronic Imaging*, Vol. 6, No. 1, pp. 54–67, January 1997.

4. E.R. Dougherty and R.P. Loce, "Optimal Mean-Absolute-Error Hit-or-Miss Filters: Morphological Representation and Estimation of the Binary Conditional Expectation," *Optical Engineering*, Vol. 32, No. 4, pp. 815–827, April 1993.
5. E.R. Dougherty and R. P. Loce, "Precision of Morphological-Representation Estimators for Translation-invariant Binary Filters: Increasing and Nonincreasing," *Signal Processing*, Vol. 40, pp. 129–154, 1994.
6. N.S.T. Hirata, E.R. Dougherty, and J. Barrera, "A Switching Algorithm for Design of Optimal Increasing Binary Filters Over Large Windows," *Pattern Recognition*, Vol. 33, No. 6, pp. 1059–1081, June 2000.
7. R. Hirata Jr., E.R. Dougherty, and J. Barrera, "Aperture Filters," *Signal Processing*, Vol. 80, No. 4, pp. 697–721, April 2000.
8. O.V. Sarca, E.R. Dougherty, and J. Astola, "Secondarily Constrained Boolean Filters," *Signal Processing*, Vol. 71, No. 3, pp. 247–263, December 1998.
9. J. Barrera, E.R. Dougherty, and M. Brun, "Hybrid human-machine binary morphological operator design. An independent constraint approach," *Signal Processing*, Vol. 80, No. 8, pp. 1469–1487, August 2000.
10. E. R. Dougherty, J. Barrera, G. Mozelle, S. Kim, and M. Brun, "Multiresolution Analysis for Optimal Binary Filters," *Mathematical Imaging and Vision*, Vol. 14, pp. 53–72, 2001.
11. O.V. Sarca, E.R. Dougherty, and J. Astola, "Two-stage Binary Filters," *Electronic Imaging*, Vol. 8, No. 3, pp. 219–232, July 1999.
12. N.S.T. Hirata, E. R. Dougherty, and J. Barrera, "Iterative Design of Morphological Binary Image Operators," *Optical Engineering*, Vol. 39, No. 12, pp. 3106–3123, December 2000.
13. E.R. Dougherty, "Optimal Mean-Square N-Observation Digital Morphological Filters I. Optimal Binary Filters," *CVGIP: Image Understanding*, Vol. 55, No. 1, pp. 36–54, January 1992.
14. E.R. Dougherty, "Optimal Mean-Square N-Observation Digital Morphological Filters II. Optimal Gray-Scale Filters," *CVGIP: Image Understanding*, Vol. 55, No. 1, pp. 55–72, January 1992.
15. R.P. Loce and E.R. Dougherty, "Optimal Morphological Restoration: The Morphological Filter Mean-Absolute-Error Theorem," *Visual Communication and Image Representation*, Vol. 3, No. 4, pp. 412–432, December 1992.
16. R.P. Loce and E.R. Dougherty, "Facilitation of Optimal Binary Morphological Filter Design Via Structuring Element Libraries and Design Constraints," *Optical Engineering*, Vol. 31, No. 5, pp. 1008–1025, May 1992.
17. R.P. Loce and E.R. Dougherty, "Mean-Absolute-Error representation and Optimization of Computational-Morphological Filters," *Graphical Models and Image Processing*, Vol. 57, No. 1, pp. 27–37, 1995.
18. E.J. Coyle and J.-H. Lin, "Stack Filters and the Mean Absolute Error Criterion," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 36, No. 8, pp. 1244–1254, August 1988.
19. J.T. Astola and P. Kuosmanen, "Representation and optimization of stack filters," In E.R. Dougherty and J.T. Astola, editors, *Nonlinear Filters for Image Processing*, pp. 237–279. SPIE and IEEE Press, Bellingham, 1999.
20. M. Gabbouj and E.J. Coyle, "Minimum Mean Absolute Error Stack Filtering with Structural Constraints and Goals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 38, No. 6, pp. 955–968, June 1990.
21. P. Kuosmanen and J.T. Astola, "Optimal stack filters under rank selection and structural constraints," *Signal Processing*, Vol. 41, pp. 309–338, 1995.
22. I. Tăbuş, D. Petrescu, and M. Gabbouj, "A training Framework for Stack and Boolean Filtering – Fast Optimal Design Procedures and Robustness Case Study," *IEEE Transactions on Image Processing*, Vol. 5, No. 6, pp. 809–826, June 1996.
23. H.J.A.M. Heijmans, *Morphological Image Operators*, Academic Press, Boston, 1994.
24. G.J.F. Banon and J. Barrera, "Minimal Representations for Translation-Invariant Set Mappings by Mathematical Morphology," *SIAM J. Applied Mathematics*, Vol. 51, No. 6, pp. 1782–1798, December 1991.
25. G.J.F. Banon and J. Barrera, "Decomposition of Mappings between Complete Lattices by Mathematical Morphology, Part I. General Lattices," *Signal Processing*, Vol. 30, pp. 299–327, 1993.
26. E.R. Dougherty, *Random Processes for Image and Signal Processing*, SPIE and IEEE Presses, Bellingham, 1999.
27. T.M. Mitchell. *Machine Learning*, "McGraw-Hill Series in Computer Science," McGraw-Hill, March 1997.
28. M. Schmitt and F. Preteux, *Image Analysis and Mathe-matical Morphology*, Vol. 2, chapter Boolean texture analysis and synthesis, Ed. J. Serra, Academic Press, New York, 1988.

Junior Barrera received the degree of doctor in automatic control and systems from the University of São Paulo in 1992. From June 1992 to 1999, he was an assistant professor at the Department of Computer Science of the Institute of Mathematics and Statistics of the University of São Paulo. Currently, he is an associate professor at the same department and he is also the scientific director of the Bioinformatic Research Center of São Paulo State University. In the last 15 years he has worked in the field of image processing analysis by mathematical morphology. He has several international publications with theoretical and applied contributions. He is co-author of the proof that any lattice operator can be decomposed in terms of elementary operators of mathematical morphology. He is also co-author of the MMach Toolbox, a very popular software for image analysis by mathematical morphology.

Roberto Hirata, Jr. received his PhD (2001) and MSc (1997) degrees in Computer Science from Universidade de SûÈo Paulo, Brazil, and he is an assistant professor at SENAC College of Computer Science and Technology since 2001. From September, 1997 to January, 1999, he worked with Dr. Edward R. Dougherty at Texas A&M, US, as part of his PhD studies. His main research interests are automatic design of morphological image operators and bioinformatics.

Marcel Brun is a PhD student at the Department of Computer Science of the Institute of Mathematics and Statistics of the University of São Paulo, Brazil. He received the bachelor degree in Mathematical Sciences from the National University of Mar del Plata, Argentina, in 1997. In 2001, he was a Visiting Research Associate in the Electrical Engineering department at Texas A&M University. His research field is in image analysis by mathematical morphology and multi-scale morphological operators design.

Edward Dougherty is a professor of Electrical Engineering at Texas A&M University. He holds a Ph.D. in mathematics from Rutgers University and an M.S. in computer science from Stevens Institute of Technology. He is an SPIE fellow, former editor of the SPIE/IS&T Journal of Electronic Imaging, author of eleven books, and editor of four collections. His current research focuses on genomic signal processing.