

# Interval + Image = Wavelet: For Image Processing under Interval Uncertainty, Wavelets are Optimal

Alejandro E. Brito and Olga Kosheleva

Department of Electrical and Computer Engineering  
University of Texas at El Paso  
El Paso, TX 79968, USA  
email {alexbr,olga}@ece.utep.edu

## Abstract

In computer and electronic manufacturing, it is very important to be able to automatically check whether the surface mounted devices (SMD) are correctly placed on the printed circuit boards. The inspection of these boards has to be done on a shop floor, where statistical characteristics of the noise vary so much that, in essence, we only have interval estimates for this noise.

We show that under this interval uncertainty, the optimal image processing technique consists of using *Haar wavelets*. Wavelets indeed lead to much better results than previously used Fourier transform techniques.

On a more fundamental level, our result is a step towards solving an important problem related to wavelets: that wavelet transforms often empirically work much better than other methods, but there are very few theoretical explanations of this efficiency. Our results shows that, probably, such a theoretical explanation can be obtained if we take interval uncertainty into consideration.

## 1 Introduction to the problem

**Case study: inspection of surface mounted devices.** Modern electronics manufacturing requires fast and efficient production. As a result, the assembly of printed circuit boards (PCB) with surface mounted devices (SMD) is usually done by robots. This manufacturing and assembly process is usually at the edge of the current manufacturing abilities, with a reasonable amount of PCB produced with defects. Therefore, it is extremely important to inspect and test the devices in order to weed out the defective ones.

Most SMD devices are so small that it is very difficult and very time-consuming for a human inspector to check whether the device is mounted at all, and whether it is mounted correctly. This problem is further complicated by the fact that many PCB have hundreds of SMD. Therefore, we need an *automatic* inspection. For that, we take a photo of the board, and we process the resulting image; see, e.g., [13, 4, 6, 7] and references therein.

**It is necessary to compress the image.** The image that we need to process is a photo. The camera produces an array of electronic signals  $\tilde{f}(x)$  that describe the brightnesses  $f(x)$  at different pixels  $x$ . The values  $\tilde{f}(x)$  corresponding to different pixels  $x$  are fed into the computer.

In the manufacturing environment, we need to process lots of images, and for each image, we need to process all these values fast, ideally, on a reasonably cheap PC-type computer. If we were to apply complicated processing techniques to *all* the pixel values, this would require lots of computer processing time: a good image consists of about 1 million pixels, and even the most crude images that we have been processing still consist of  $53 \times 27 \approx 1,400$  pixels. It is therefore desirable to *compress* this data to a few numbers, and then base our decisions on the compressed data only.

**How can we compress the image?** How is data compressed in general? For example, how do physicists represent a dependency  $y = f(x)$  between the two quantities?

Usually, this dependency is smooth (and even analytical), and therefore, the corresponding function  $f(x)$  can be represented (at least for small  $x$ ) as a sum of its Taylor series:

$$f(x) = f(0) + f'(0) \cdot x + \frac{1}{2} \cdot f''(0) \cdot x^2 + \dots$$

Since measurements are usually imprecise, we do not need all these terms to represent the measurement results; it is sufficient to take only a few first terms in this expansion:

- To describe the most crude measurements, it may be possible to keep only the first (constant) term and take  $f(x) \approx c_1$  for some constant  $c_1$ .
- To describe better measurements, we may need linear terms as well. In other words, we take  $f(x) \approx c_1 + c_2 \cdot x$  and use the coefficients  $c_1$  and  $c_2$  to make decisions about the analyzed dependence.
- To get an even better approximation, we may want to retain quadratic terms as well, and take  $f(x) \approx c_1 + c_2 \cdot x + c_3 \cdot x^2$ , etc.

In this case, we start with a *basis* consisting of the functions  $e_1(x) = 1$ ,  $e_2(x) = x$ ,  $e_3(x) = x^2$ , etc., a basis in which every function (at least every function that is smooth enough) can be represented as an infinite series

$$f(x) = c_1 \cdot e_1(x) + c_2 \cdot e_2(x) + \dots + c_k \cdot e_k(x) + \dots, \quad (1)$$

and then we take several first coefficients  $c_1, \dots, c_N$  as the desired compressed representation of the function  $f(x)$ .

In most physical cases, the monomials  $1, x, x^2$ , etc., form a physically reasonable basis; in other cases, sines or other functions may be a better first approximation than linear or quadratic ones. But in general, the idea of using the first few coefficients of the expansion seems to be a reasonable data compression method.

This approach is used in *imaging* as well. In this paper, therefore, we will consider data compression methods based on this idea.

It is important to mention that in imaging problems, usually, the image is 2-dimensional. The corresponding function  $f(x)$  is a function of *two* variables (i.e.,  $x = (x_1, x_2)$ ), and therefore, we must consider infinite series (1) in which the basis functions  $e_i(x) = e_i(x_1, x_2)$  are also functions of *two* variables.

*Comments.*

- From the mathematical viewpoint, this *compression* of a function  $f(x)$  is the same as taking a *projection* of this function  $f(x)$  onto a suitable finite-dimensional function space, a space that is defined by finitely many basis elements.
- In real life, observations are limited to a bounded 1D or 2D area. Therefore, from the mathematical viewpoint, these images are represented by functions defined on an interval (in a 1D case) or on a bounded area (in 2D case), i.e., functions that are equal to 0 outside this area. Some image processing methods only use functions defined on this same interval (bounded area); however, there are many other image efficient processing techniques, e.g., methods based on Fourier transform, that use functions different from 0 *outside* this interval (bounded area) as well. In view of this possibility, we will not restrict ourselves *a priori* only to functions that are equal to 0 outside the given area. (Not surprisingly, however, this restriction will automatically appear when we find the *optimal* image processing technique for our problem.)

**It is reasonable to use orthonormal bases.** Some of the known bases (e.g., sines and cosines) are *orthonormal* in the sense that

$$\int e_i(x) \cdot e_j(x) dx = 0 \text{ when } i \neq j; \quad (2)$$

$$\int e_i^2(x) dx = 1 \text{ for all } i. \quad (3)$$

In principle, a physical basis need not be orthonormal; e.g., the monomials do not have this property. However, it is always possible to transform each basis  $\{e_1(x), e_2(x), \dots\}$  of functions into a new orthonormal basis  $\{\bar{e}_1(x), \bar{e}_2(x), \dots\}$ :

in essence, this transformation can be performed by using the known orthonormalization procedure:

$$\begin{aligned}\bar{e}_1(x) &= g_{11} \cdot e_1(x); \\ \bar{e}_2(x) &= g_{21} \cdot e_1(x) + g_{22} \cdot e_2(x);\end{aligned}$$

etc., for appropriate coefficients  $g_{ij}$ ; e.g.,  $g_{11} = 1/\sqrt{\int e_1^2(y) dy}$ .

*Comment.* To be more precise, there may be problems with this procedure for an *infinite-dimensional* function space, but when we represent data inside the computer, we always restrict the number  $N$  of coefficients that we can store; thus, in effect, we always consider a *finite-dimensional* space of functions of the type  $c_1 \cdot e_1(x) + \dots + c_N \cdot e_N(x)$ . For bases in such finite-dimensional spaces, the above procedure works perfectly well.

When we move to an orthonormal base, we, in general, *do not* seem to *lose* anything, and we do not change the actual compression: indeed, for all  $N$ , the class of all functions of the type  $c_1 \cdot e_1(x) + \dots + c_N \cdot e_N(x)$  is exactly the same as the class of all functions of the type  $c_1 \cdot \bar{e}_1(x) + \dots + c_N \cdot \bar{e}_N(x)$ . On the other hand, we do *gain* in computation time when we turn to orthonormal bases (and computational time is what we try to minimize in the first place): Indeed, for orthonormal bases, the computation of the coefficients becomes much computationally simpler than for the general bases:

$$c_i = \int f(y) \cdot \bar{e}_i(y) dy. \quad (4)$$

In view of this important advantage of orthonormal bases (and also in view of the absence of any disadvantages), it makes perfect sense to use such bases.

In the following text, we will therefore assume that the basis  $e_i(x)$  is orthonormal, i.e., that the conditions (2)–(3) are satisfied.

*Comment.* The fact that by an orthonormalization, we do not seem to lose anything *in general* does not mean that in specific cases, using non-orthonormal bases is *always* worse than using their orthonormalizations. Sometimes the original functions  $e_i(x)$  are easy-to-compute, while the functions  $\bar{e}_i(x)$  from the orthonormalized basis require more computations.

In this case, i.e., when the orthonormal basis is computationally complicated, it may be better to go back to the original non-orthonormal basis. By going back to that original basis:

- we lose some computation time on computing the coefficients, but
- we also gain computation time on computing the values of the basis functions.

This going back has indeed been shown to be very beneficial in different areas of data processing, including image processing (see, e.g., the success of a special non-orthonormal basis called bi-orthogonal wavelets, [12], Section 4.7).

So:

- If the orthonormal basis turns out to be computationally complicated, it makes sense to consider possible non-orthonormal alternatives.
- In *our problem*, however, we will see that the optimal orthonormal basis consists of the functions that are extremely simple to compute. Therefore, for our problem, there seems to be no reason to consider non-orthonormal bases.

**Taking measurement errors into consideration.** Brightness values are measured with a non-negligible inaccuracy. In general, in image processing, we use *statistical* methods of processing data that are based on the assumption that we know the probabilities of different imaging errors. It is often possible to collect these probabilities: we measure the frequencies of different errors and we see that as the number of experiments grow, these frequencies tend to a certain limit which is the desired probability.

However, in our case, we are dealing with images taken on the shop floor. The situation on the shop floor changes so frequently and so unpredictably that there are no stable frequencies of errors. Therefore, we do not know the probabilities of different value of noise; the only information that we have about the noise is the upper bound  $\Delta$  on its value: For every point  $x$ , the difference  $\Delta f(x) = f(x) - \tilde{f}(x)$  between the actual (unknown) brightness  $f(x)$  and its measured value  $\tilde{f}(x)$  is bounded by  $\Delta$ :

$$|\Delta f(x)| \leq \Delta.$$

In other words, after measuring brightness, we only know, for each point  $x$ , the *interval*  $[\tilde{f}(x) - \Delta, \tilde{f}(x) + \Delta]$  of possible values of brightness  $f(x)$  at this point  $x$ .

This measurement inaccuracy leads to inaccuracy in the coefficients  $c_i$ , i.e., to the difference  $\Delta c_i = c_i - \tilde{c}_i$  between the *ideal* values (4) of these coefficients (the values that we would have gotten if we had the ideal image) and the actually *computed* values

$$\tilde{c}_i = \int \tilde{f}(y) \cdot e_i(y) dy. \quad (5)$$

If this inaccuracy is huge, then the resulting values of the coefficients  $c_i$  are very unreliable and cannot be used to make any conclusions about the actual image. So, we must make this difference as small as possible.

How can we express this idea numerically? If we reconstruct the image from the compressed data, we get the following formula:

$$\tilde{f}_{\text{rec}}(x) = \tilde{m}_1(x) + \dots + \tilde{m}_N(x),$$

where we denoted

$$\tilde{m}_i(x) = \tilde{c}_i \cdot e_i(x). \quad (6)$$

If we used the precise image instead, we would have gotten a similar (but more accurate) representation

$$f_{\text{rec}}(x) = m_1(x) + \dots + m_N(x),$$

where we denoted

$$m_i(x) = c_i \cdot e_i(x). \quad (7)$$

It makes sense, therefore, to estimate the relative *quality* of choosing a function  $e_i(x)$  as the largest possible value of the difference

$$\Delta m_i(x) = m_i(x) - \tilde{m}_i(x). \quad (8)$$

So, we arrive at the following definitions.

## 2 Mathematical formulation of the problem

Let  $X$  be a space with a measure  $\mu$  (e.g., a line  $R$  or a plane  $R^2$  equipped with a standard (Lebesgue) measure  $\mu$ ). We will consider real-valued functions  $f : X \rightarrow R$  that are defined on the entire space  $X$  (in other words, we consider *point-wise defined* function).

As usual, we say that a function  $f : X \rightarrow R$  is *square integrable* if it is measurable with respect to measure  $\mu$  and if the integral of its square is finite, i.e.,  $\int f(x)^2 dx < +\infty$ .

*Comment.* To prevent possible misunderstanding, we want to emphasize the following.

The set of all square integrable functions is closely related to the well-known function space  $L^2(X)$ : namely,  $L^2(X)$  is formed by equivalence classes of square integrable functions with respect to equivalent relation  $f \sim g$  if and only if  $f(x) = g(x)$  for almost all  $x \in X$  (i.e., for all  $x \in X$  except maybe for a set of measure 0). Usually, people speak interchangeably about the pointwise defined functions and the corresponding equivalent classes from  $L^2$ , and it causes no confusion as long as we talk about the properties that change a function to an equivalent one.

Some of the properties in which we are interested are just like that: e.g., the property that two functions  $f_1(x)$  and  $f_2(x)$  are orthogonal to each other ( $\int f_1(x) \cdot f_2(x) dx = 0$ ) does not change if we replace  $f_1(x)$  and  $f_2(x)$  by equivalent functions  $g_1 \sim f_1$  and  $g_2 \sim f_2$ .

However, other characteristics in which we are interested, such as the supremum  $\sup f(x)$  of the function, may change drastically if we replace the original function  $f$  by an equivalent function  $g \sim f$ . In this case, it is important to consider pointwise defined functions rather than their equivalent classes (i.e., elements of  $L^2(X)$ ).<sup>1</sup>

---

<sup>1</sup>We are thankful to the anonymous referee who attracted our attention to this important potential source of confusion.

By an *orthonormal basis*, we mean a sequence of square integrable functions  $e_i : X \rightarrow \mathbb{R}$  that satisfy the properties (2)–(3), and which generates the entire space  $L^2(X)$  (i.e., for which every function from the set  $L^2(X)$  can be represented as a  $L^2$ -limit of linear combinations of functions from this basis).

**Definition 1.** By a *compression scheme*, we mean a pair  $(\{e_1(x), e_2(x), \dots\}, N)$ , in which:

- $\{e_1(x), e_2(x), \dots\}$  is an orthonormal basis, and
- $N$  is an integer.

**Definition 2.** Let  $(\{e_i(x)\}, N)$  be a compression scheme, and let  $\Delta > 0$  be a positive real number. This number  $\Delta$  will be called *measurement inaccuracy*. By the *reconstruction inaccuracy*  $q_i$  of  $i$ -th term of this basis  $e_i(x)$ , we mean the largest possible value  $q_i$  of the difference  $|\Delta m_i(x)|$ :

$$q_i = \sup_{x, f(x), \Delta f(x)} |\Delta m_i(x)|,$$

where  $\Delta m_i(x)$  is determined by the formulas (4)–(8) with  $\tilde{f}(x) = f(x) - \Delta f(x)$ , and supremum is taken over:

- all points  $x \in X$ ,
- all square integrable functions  $f$ ,
- all square integrable functions  $\Delta f$  for which  $|\Delta f(x)| \leq \Delta$  for all  $x \in X$ .

*Comment.* In general, the supremum (least upper bound) is defined for an arbitrary set of real numbers; for some sets (e.g., for the entire set of real numbers  $\mathbb{R}$ ) it can be infinite. Therefore, the supremum  $q_i$  of all the values  $|\Delta m_i(x)|$  for all possible  $x \in X$ , and  $f$  and  $\Delta f$  as described in Definition 2 is defined for all possible compression schemes; for some compression schemes, this supremum is finite; for some other schemes, it may be infinite.

Our goal is to find, for each  $N$ , the basis for which the reconstruction inaccuracy is the smallest possible. We want to minimize  $N$  different numbers  $q_1, \dots, q_N$ . In general, if we minimize one objective function, it is difficult to expect that any other objective function will be simultaneously minimized. However, in our case, we are lucky: for the cases of  $X = \mathbb{R}$  and  $X = \mathbb{R}^2$  that correspond to imaging, there is a basis for which *all* the reconstruction inaccuracies  $q_i$  take the smallest possible value.

*Comment.* As we have mentioned, for some compression schemes, some of the values  $q_i$  can be infinite. Since we are looking for a compression scheme with the smallest possible values of  $q_i$ , we, thus, are interested in the compression schemes for which all the reconstruction inaccuracies  $q_i$  are finite.

**Definition 3.** Let  $\phi(x)$  be a function that is equal to 1 for  $0 \leq x < 1$  and to 0 otherwise. Then, the function  $w(x) = \phi(2 \cdot x) - \phi(2 \cdot x - 1)$  is equal:

- to 1 for  $0 \leq x < 1/2$ ,
- to  $-1$  for  $1/2 \leq x < 1$ , and
- to 0 for all other  $x$ .

By a 1D Haar basis, we mean the basis that consists of the function  $\phi(x)$  and of the functions  $w_{jk}(x) = 2^{j/2} \cdot w(2^j \cdot x - k)$ , where  $j$  and  $k$  are arbitrary integers.

*Comments.* In general, if we have a function  $w(x)$  that tends to 0 as  $|x| \rightarrow \infty$ , and for which the functions  $w(2^j \cdot x - k)$  are *orthogonal* to each other, so that after normalization, we get an *orthonormal* basis  $c_{jk} \cdot w(2^j \cdot x - k)$ , then this basis is called an (orthonormal) *wavelet basis*. We need to add one or several additional functions  $\phi_a(x)$  (their number and type differ for different functions  $w(x)$ ) to it to make it a (complete) basis for the entire space  $L^2(\mathbb{R})$ ; see, e.g., [2, 8, 12, 15].

The reason why we need an extra function (or functions)  $\phi_a(x)$  can be easily illustrated on the example of the Haar wavelet basis: as one can easily show, for the corresponding function  $w(x)$  (from Definition 3), we have  $\int w(x) dx = 0$  and therefore  $\int w_{jk}(x) dx = 0$  for all  $j$  and  $k$ . Hence, for an arbitrary linear combination  $f(x) = \sum a_{jk} \cdot w_{jk}(x)$  of these functions, we also have  $\int f(x) dx = 0$ . This means, in particular, that we cannot represent a function  $f(x)$  with  $\int f(x) dx \neq 0$  as the desired linear combination, i.e., that the functions  $w_{jk}(x)$  do not form a *complete* basis. It turns out that this basis can be completed if we add the function  $\phi(x)$  as described in Definition 3 (see, e.g., [12], Sections 2.2. and 2.10).

**Definition 4.** By a 2D Haar basis, we mean the basis consisting of the functions  $f_{ij}(x_1, x_2) = e_i(x_1) \cdot e_j(x_2)$ , where  $e_i$  and  $e_j$  are functions from the 1D Haar basis.

**THEOREM.** For every  $i$ , the imaging inaccuracy of the Haar wavelet basis is the smallest possible.

*Comments.*

- The proof of this theorem is given in the following section.
- For SMD, wavelets, in particular, Haar wavelets, indeed lead to much better results than more traditional Fourier transform techniques [9, 10, 6, 1, 7].
- On a more *fundamental* level, our result is a step towards solving the following important problem related to wavelets:

- wavelet transforms *often empirically* work much *better* than other methods, but
- there are very *few theoretical explanations* of this efficiency.

Our result shows that, at least on some cases, such a theoretical explanation can be obtained if we take interval uncertainty into consideration.

- A related problem is: *Which wavelet is the best?* This problem is raised, e.g., in [12].
  - In [11] and [3], this problem is analyzed under the assumption that the measurement errors are *random*, Gaussian, and independent. In this case, the best approximation corresponds to minimizing the sum of the squares of these errors. Special wavelets are presented that minimize this sum.
  - In our paper, we consider a similar problem, but under the assumption that the measurement errors belong to the corresponding *intervals*. The optimization of the corresponding worst-case error leads to Haar wavelets.

### 3 Proof

1. Let us first show that for an arbitrary basis  $\{e_1(x), e_2(x), \dots\}$ , we have

$$q_i = \Delta \cdot \sup_x |e_i(x)| \cdot \int |e_i(y)| dy.$$

Indeed, for each function  $f(x)$ , we have

$$\begin{aligned} \Delta m_i(x) &= m_i(x) - \tilde{m}_i(x) = c_i \cdot e_i(x) - \tilde{c}_i \cdot e_i(x) = \\ &= e_i(x) \cdot \int f(y) \cdot e_i(y) dy - e_i(x) \cdot \int \tilde{f}(y) \cdot e_i(y) dy = \\ &= e_i(x) \cdot \int e_i(y) \cdot \Delta f(y) dy. \end{aligned}$$

Hence,  $|\Delta m_i(x)| = |I| \cdot |e_i(x)|$ , where we denoted  $I = \int e_i(y) \cdot \Delta f(y) dy$ . Since the function  $|\Delta m_i(x)|$  is proportional to  $|e_i(x)|$ , the supremum of  $|\Delta m_i(x)|$  is proportional to the supremum of  $|e_i(x)|$ , i.e.,

$$\sup_x |\Delta m_i(x)| = |I| \cdot \sup_x |e_i(x)|.$$

The larger  $|I|$ , the larger  $\sup_x |\Delta m_i(x)|$ . Therefore, the supremum  $q_i$  of the value  $\sup_x |\Delta m_i(x)|$  is equal to

$$q_i = \sup_{x, f, \Delta f} |\Delta m_i(x)| = \sup_{\Delta f} |I| \cdot \sup_x |e_i(x)|,$$

where the supremum of  $|I|$  is taken over all square integrable functions  $\Delta f$  for which  $|\Delta f(x)| \leq \Delta$  for all  $x \in X$ . So, to compute  $q_i$ , we must compute the supremum of  $|I|$ .

To compute this supremum, let us first find the upper estimate for  $|I|$ . Since  $I = \int e_i(y) \cdot \Delta f(y) dy$ , we would like to use the inequality

$$|I| = \left| \int e_i(y) \cdot \Delta f(y) dy \right| \leq \int |e_i(y) \cdot \Delta f(y)| dy.$$

However, this inequality only holds when its right-hand side is defined. Both functions  $e_i(y)$  and  $\Delta f(y)$  are square integrable and therefore, measurable. Therefore, their product  $e_i(y) \cdot \Delta f(y)$  is also a measurable function, and hence, so is the absolute value  $|e_i(y) \cdot \Delta f(y)|$  of this product. This absolute value is, therefore, a non-negative measurable function, and it is known that every such function has an integral (although this integral may be infinite). Thus, we can use the above inequality.

Similarly, we conclude that

$$\begin{aligned} |I| &= \left| \int e_i(y) \cdot \Delta f(y) dy \right| \leq \int |e_i(y) \cdot \Delta f(y)| dy \leq \\ &\int \Delta \cdot |e_i(y)| dy = \Delta \cdot \int |e_i(y)| dy. \end{aligned}$$

(The existence of the integral  $\int |e_i(y)| dy$  also follows from the fact that  $|e_i(y)|$  is a non-negative measurable function; in general, this integral can be infinite.)

On the other hand, for arbitrary  $B$ , we can take  $\Delta f(y) = \Delta \cdot \text{sign}(e_i(y))$  for  $|y| \leq B$  and  $\Delta f(y) = 0$  otherwise. For this choice of  $\Delta f(y)$ , we have

$$I = \int e_i(y) \cdot \Delta f(y) dy = \Delta \cdot \int_{|y| \leq B} |e_i(y)| dy.$$

The larger  $B$ , the closer this value to  $\Delta \cdot \int |e_i(y)| dy$ ; in more precise terms,

$$\int |e_i(y)| dy = \sup_B \int_{|y| \leq B} |e_i(y)| dy.$$

Therefore, the supremum of possible values of  $|I|$  is indeed  $\Delta \cdot \int |e_i(y)| dy$ , and therefore, the supremum value  $q_i$  of values  $|\Delta m_i(x)|$  is indeed equal to

$$\Delta \cdot \sup_x |e_i(x)| \cdot \int |e_i(y)| dy.$$

The formula is proven.

2. Let us now show that for an arbitrary basis, we have  $q_i \geq \Delta$ .

Indeed, clearly, for all  $y$ ,

$$\sup_x |e_i(x)| \geq |e_i(y)|.$$

Therefore,

$$\begin{aligned} \frac{q_i}{\Delta} &= \sup_x |e_i(x)| \cdot \int |e_i(y)| dy = \\ &= \int \sup_x |e_i(x)| \cdot |e_i(y)| dy \geq \int |e_i(y)| \cdot |e_i(y)| dy. \end{aligned}$$

Since  $\{e_i(x)\}$  is an orthonormal basis, the right-hand side is equal to 1 and therefore,  $q_i/\Delta \geq 1$  and  $q_i \geq \Delta$ .

*Comment.* In essence, in this part of the proof, we are proving and using, for  $f(x) = e_i(x)$ , the known Hölder's inequality  $\|f\|_{L^2}^2 \leq \|f\|_{L^1} \cdot \|f\|_{L^\infty}$  (see, e.g., [5]), where  $\|f\|_{L^2} = \sqrt{\int |f(x)|^2 dx}$ ,  $\|f\|_{L^1} = \int |f(x)| dx$ , and  $\|f\|_{L^\infty} = \sup\{|f(x)|\}$ .

3. To complete the proof, we must show that for Haar basis,  $q_i = \Delta$ .

Indeed, for each of the functions from this basis, there exists a non-zero real number  $C$  such that for every  $x$ , either  $e_i(x) = 0$ , or  $|e_i(x)| = C$ . Therefore,

$$\frac{q_i}{\Delta} = \sup_x |e_i(x)| \cdot \int |e_i(y)| dy = C \cdot \int |e_i(y)| dy = \int C \cdot |e_i(y)| dy.$$

We know that for every  $y$ , either  $e_i(y) = 0$ , or  $|e_i(y)| = C$ . Therefore:

- When  $e_i(y) = 0$ , we have  $C \cdot |e_i(y)| = 0 = e_i^2(y)$ .
- When  $|e_i(y)| = C$ , then  $C \cdot |e_i(y)| = e_i^2(y)$ .

In both cases,  $C \cdot |e_i(y)| = |e_i(y)|^2$  and therefore,  $q_i/\Delta = \int C \cdot |e_i(y)| dy = \int |e_i(y)|^2 dy$ . Since  $\{e_i(y)\}$  is an orthonormal basis, we have  $q_i/\Delta = 1$  and  $q_i = \Delta$ .

*Comment.* In essence, we have shown that the functions  $e_i(x)$  from the Haar basis satisfy the equality  $\|e_i\|_{L^2}^2 = \|e_i\|_{L^1} \cdot \|e_i\|_{L^\infty}$ .

The statement is proven, and so is the theorem.

**Acknowledgments.** This paper was partially supported by NSF grant CDA-9522207.

The authors are thankful to Sergio Cabrera for his supervision, and to Vladik Kreinovich and Mikhail Zakharevich for valuable discussions. We are also very thankful to the anonymous referees for important suggestions, and to Günter Mayer, the editor of the student issue, for his help.

## References

- [1] G. Carrillo, S. D. Cabrera, and A. A. Portillo, "Inspection of Surface-Mount-Device images using wavelet processing", *Proceedings of SPIE Applied Imagery Pattern Recognition*, Washington D.C., October 1996.

- [2] C. Chui, *An Introduction to Wavelets*, Academic Press, 1992.
- [3] R. R. Coifman, G. Matviyenko, and Y. Meyer, ‘Modulated Malvar-Wilson bases’, *Applied and Computational Harmonic Analysis*, 1997, Vol. 4, pp. 58–61.
- [4] *Computer vision: theory and industrial applications*, Springer-Verlag, Berlin-N.Y., 1992.
- [5] R. E. Edwards, *Functional analysis. Theory and applications*, Dover, N.Y., 1995.
- [6] J. M. Gallegos, J. R. Villalobos, G. Carrillo, and S. D. Cabrera, “Sequential algorithms for the inspection of Surface Mounted Devices”, *Proceedings of the 13th SPIE Conference on Intelligent Robots and Computer Vision, Boston, MA, SPIE*, 1994, Vol. 2354, pp. 215–226.
- [7] J. M. Gallegos, J. R. Villalobos, G. Carrillo, and S. D. Cabrera, “Reduced-dimension and wavelet processing of SMD images for real-time inspection”, *Proceedings of the IEEE Southwest Symposium on Images Analysis and Interpretation*, San Antonio, April 1996.
- [8] S. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, Vol. 14, No. 7, pp. 710–732.
- [9] S. Mallat and W. L. Hwang, “Singularity detection and processing with wavelets”, *IEEE Transactions on Information Theory*, 1992, Vol. 38, No. 2, pp. 617–643.
- [10] S. Mallat and W. L. Hwang, “Characterization of signals from multiscale edges”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, Vol. 14, No. 7, pp. 710–732.
- [11] G. Matviyenko, “Optimized local trigonometric bases”, *Applied and Computational Harmonic Analysis*, 1996, Vol. 3, pp. 301–323.
- [12] Y. Meyer, *Wavelets algorithms and applications*, SIAM, Philadelphia, 1993.
- [13] L. F. Pau, *Computer vision for electronics manufacturing*, Plenum, N.Y., 1990.
- [14] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA, 1996.
- [15] M. Vetterli and J. Kovacevic, *Wavelets and subband coding*, Prentice Hall, Englewood Cliffs, New Jersey, 1995.