Visual Shape Analysis by Wavelets

Evangelia Micheli-Tzanakou and Ivan Marsic
Department of Biomedical Engineering, Rutgers University
Piscataway, NJ 08855-0909

Abstract. Recent complexity analysis shows that the problem of visual recognition is an intractable problem, as is the training of neural networks. We propose a representation scheme which greatly simplifies recognition. The approach relies on the wavelet analysis, a recent tool developed for local space–frequency analysis of a broad class of functions.

Based on two assumptions, (*i*) visual objects are localized in the space domain, and (*ii*) the recognition system is robust, we conclude that for the purpose of recognition every object *is localized in the frequency domain* as well. This fact, combined with the properties of wavelet transform, allows us to keep the same amount of information for objects of any size.

We further analyze the consequences of spatial and frequency localization of the objects, and a new framework for visual object representation and recognition arises from this analysis.

1 Introduction

All recent work on complexity analysis of tasks that are faced by the biological information processing systems indicates that these tasks are intractable in the sense of computational complexity theory [4, 7]. Yet, biological systems solve most of these tasks without any apparent problems. Which simplifications are used to manage the inherent intractability of these tasks? There are many suggestions about the possible strategies used [4, 7]. The purpose of this paper is to investigate the inherent properties of visual objects and to propose a representation scheme which will greatly simplify the recognition task.

Tsotsos [7] has considered a visual search task, the experimental paradigm of which is as follows: Given a set of memory items or targets and a test display that contains several non–target items and may or may not contain targets, measure the length of time a subject needs to detect a given number of the targets in the display. It is a categorization task in which a subject must distinguish between at least two class of signals: goal signals which must be located and reported and background signals which must be ignored. In its full generality it includes the possibility of noisy or partial matches. The task has two variants: (a) *unbounded visual search*, in which either the target is explicitly unknown in advance or it is somehow not used in the execution of the search, and (b) *bounded visual search*, in which the target is explicitly known in advance in some form that enables explicit bounds to be determined that can be used to limit the search process.

Tsotsos has shown that unbounded visual search is inherently NP–complete[1] in the size of the image. This result is an independent of implementation, that is, whether one is considering the brain, a machine, or some yet to be discovered method of implementation, the inherent complexity of the general problem remains the same and an implementation must deal with it. He defines the problem in terms of the algorithms. The solution to a visual search problem involves solving a subproblem of *visual matching*: Given a test image, a difference function, and a correlation function, is there a subset of pixels of the test image such that the value of the difference function for that subset is less that a given threshold and such that the value of correlation exceeds some

[1] The class of P–complete problems consists of all those problems that can be solved in polynomial time. In addition to the class P, which are tractable problems, there is a major class of presumably intractable or NP–complete problems.

other threshold? Tsotsos shows that the unbounded visual matching problem has exactly the same structure as a known NP–complete problem, namely the Knapsack Problem [7]. It follows that the unbounded visual search is NP–complete.

As Tsotsos states, visual search may be a very basic problem that is found in most other types of visual information processing, and especially in visual object recognition. Therefore, its complexity is a kind of lower bound for complexity of other visual tasks; general object recognition must therefore also be NP–complete.

Other difficulties with the complexity of visual object recognition arise from the analysis of the time for learning in neural networks. Judd [4] has studied the computations required to support simple associative memory (this kind of memory obviously must be present in a visual system). He has shown that loading of feedforward neural networks (training a network to remember associated pairs) is NP–complete in number of neurons. For all reasonable sets of neuron functions, his results about the difficulty of the loading problem are completely independent of the choice of type of functions that each neuron can perform.

In order to solve the intractable problem of recognition, the visual system must apply some simplifications. The question is: Is it possible to know a priori if any particular part of the available information will not be relevant to the problem? The intuitive and the formal justification that the answer in the case of visual object recognition is yes are given below.

The intuitive argument is as follows: Let us assume that there exist two visual systems with different resolutions of retinal images. For example, System 1 has 512×512 points, and System 2 has 32×32 points. Is System 2 handicapped relative to System 1? —Theoretically speaking, it is not. System 2 might obtain the same visual information about the environment as System 1 does, just by moving closer to the object of interest and by eye movements.

Are there any new problems which System 2 has to face, which are not faced by System 1? —The answer is again no. Both systems have to relate successive scenes during approaching/departing and during eye movements.

The problem for System 2 is that it has to approach an object which might be dangerous; furthermore it is time consuming to walk around it. On the other hand, it is benefited in the sense that it has to process much less input data at any time, and therefore it does not face the complexity problem.

Now, let us imagine that we have another system, System 3, which consists of many small retinas, all of which are focused at different positions and distances in space. If System 3 has an effective schedule for handling these inputs one by one, then it features benefits from both previous systems, i.e. high resolution and small amounts of input data. In fact, System 3 "normalizes" all objects to the same size (e.g. 32×32 points), and then processes them further. This normalization has to be justified, and we have to find the optimal way to do it.

The objective of this work is to show that this intuition may be supported by the wavelet analysis. This paper shows the feasibility of the approach for any visual system, though it does not argue that this path must be taken by any system. Somewhat different arguments were presented in our previous work [6]. Here we give a theoretical justification and quantitative analysis of the ideas presented therein.

2 Wavelets and Time–Frequency Localization

The approach relies on the wavelet analysis, a recent tool developed for local time(space)–frequency analysis of a broad classes of functions [1, 2, 5]. Here we will give a brief overview of definitions and some results about the wavelet transform. We follow the exposition given in [1, 2], where an

extensive treatment of wavelets can be found. Due to the simplicity, the review is restricted to one dimensional wavelets, though the similar analysis can be carried over to the two dimensional case.

To calculate the wavelet transform of a square integrable function $f \in L^2(\mathbb{R})$, we have to have a set of analyzing functions, called *wavelets*, onto which we will project the function, and so calculate the wavelet coefficients. The "mother wavelet" $\psi \in L^2(\mathbb{R})$, which generates the others, should satisfy the admissibility condition

$$C_\psi = 2\pi \int_{-\infty}^{+\infty} \frac{d\xi}{|\xi|} |\hat{\psi}(\xi)|^2 < \infty. \tag{1}$$

where $\hat{\psi}$ is the Fourier transform of the wavelet ψ. This condition implies that $\hat{\psi}(0) = 0$, and that $\hat{\psi}(\xi)$ is small enough in the vicinity of $\xi = 0$. For all practical purposes, Eq.(1) is equivalent to the requirement that it has a zero mean $\int dx \psi(x) = 0$. An example of a mother wavelet is the so called Mexican hat function, which is the second derivative of a Gaussian. We generate a doubly–indexed family of wavelets from ψ by dilating and translating,

$$\psi^{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right),$$

where $a, b \in \mathbb{R}, a \neq 0$. Scaling parameter a is related to the dilation or contraction of the mother wavelet, whereas parameter b determines its translation. In fact, scaling corresponds to the change in frequencies, as we shall see later, and translation localizes time or position.

In most of the applications, the function will be given in the discrete form, $f \in \ell^2(\mathbb{Z})$. Therefore, we have to discretize the wavelet family as well. The following discretization of wavelet parameters is usually chosen: $a = a_0^m, b = nb_0a_0^m$; $m, n \in \mathbb{Z}$. Elementary dilation step $a_0 > 1$, and elementary translation step $b_0 > 0$ are fixed. The appropriate choices for a_0, b_0 depend on the wavelet ψ. Now, the discrete family of dilated and shifted wavelets corresponds to

$$\psi_{m,n}(k) = \frac{1}{\sqrt{a_0^m}} \psi\left(\frac{k - nb_0a_0^m}{a_0^m}\right) = a_0^{-m/2} \psi(a_0^{-m}k - nb_0), \tag{2}$$

The discrete wavelet transform of function f is given by the inner products of the function with the family of wavelets

$$(T^{wav}f)(m,n) = \langle f, \psi_{m,n} \rangle = a_0^{-m/2} \sum_{k=-\infty}^{+\infty} f(k) \overline{\psi(a_0^{-m}k - nb_0)}. \tag{3}$$

where the overline denotes the complex conjugate. It can also be viewed as a convolution product $(T^{wav}f)(m,n) = f * \overline{\psi_{m,n}}$, or as a filtering of f with a band–pass filter whose impulse response is $\overline{\psi_{m,n}}$.

Instead of the admissibility condition (1), in the discrete case the mother wavelet is admissible if the $\{\psi_{m,n}; m, n \in \mathbb{Z}\}$ constitute a *frame*, i.e. there should exist $A > 0, B < \infty$ so that

$$A \| f \|^2 \leq \sum_{m,n} |\langle f, \psi_{m,n} \rangle|^2 \leq B \| f \|^2, \tag{4}$$

for all $f \in \ell^2(\mathbb{Z})$.

To reconstruct the original function, we have to determine the dual frame, defined by $\widetilde{\psi_{m,n}} = (F^*F)^{-1}\psi_{m,n}$, where $F^*F f = \sum_{m,n} \langle f, \psi_{m,n} \rangle \psi_{m,n}$. The original function can be reconstructed from wavelet coefficients using the following formula

$$f = \sum_{m,n} \langle f, \psi_{m,n} \rangle \widetilde{\psi_{m,n}} = \sum_{m,n} \left\langle f, \widetilde{\psi_{m,n}} \right\rangle \psi_{m,n}.$$

The frames, even the tight ones (for which $A = B$), are generally not (orthonormal) bases because $\psi_{m,n}$ are typically not linearly independent—a frame contains "too many" vectors. This means that for a given f, there exist many different superpositions of the $\psi_{m,n}$ which all add up to f. Using the dual frame gives the most "economical" case (see [2] for further discussion).

In the case when wavelets make orthonormal basis, even though the Nyquist sampling criterion is violated, it is possible to obtain perfect reconstruction of the original function. This is because the aliasing errors from all of the sub-bands cancel when the bands are recombined [2].

One of the main motivations for studying wavelet transforms is that they provide good time(space)–frequency localization. If the wavelet ψ is well localized both in time and frequency, then the frame generated by ψ will share that property as well.

Let us assume that $|\psi|$ and $|\hat{\psi}|$ are symetric, although it is not necessary. A convenient way to represent the wavelets is by "coherent states" generated from a single function ψ by translating in a "phase space." Phase space, a term borrowed from physics, stands for the 2–D time–frequency space, considered as one geometric whole. Then ψ is centered around 0 in time and near $\pm\xi_0$ in frequency. If ψ is well localized in time and frequency, then $\psi_{m,n}$ will similarly be well localized around $a_0^m n b_0$ in time and around $\pm a_0^{-m}\xi_0$ in frequency.

We will only concern ourselves with the first quadrant of the phase space (positive time and positive frequencies). The point around which the mother wavelet ψ is concentrated is given by the first moment or mean value

$$\begin{aligned} x_0 &= \frac{\int_{-\infty}^{\infty} dx\, x|\psi(x)|^2}{\int_{-\infty}^{\infty} dx|\psi(x)|^2}, \\ \xi_0 &= \frac{\int_0^{\infty} d\xi\, \xi|\hat{\psi}(\xi)|^2}{\int_0^{\infty} d\xi|\hat{\psi}(\xi)|^2}, \end{aligned} \tag{5}$$

For any wavelet $\psi_{m,n}$ of the family, the center is given by $< x >= a_0^m n b_0 + x_0$, and $< \xi >= a_0^{-m}\xi_0$, and these coordinates determine the lattice points in Figure 1. (Definition of a central point other than by moments is possible.) Both denominators represent the energy of ψ. The second formula relates the frequency and scale by giving the mean frequency $< \xi >$ of the wavelet $\psi_{m,n}$ of scale a_0^m.

Each wavelet is concentrated in some time–frequency window of the phase space. The whole family, obtained by dilating and translating, covers the entire phase space. The size of the mother wavelet's cell is determined by first central moments or standard deviations in time and frequency

$$\begin{aligned} \sigma_{x_0}^2 &= \frac{\int_{-\infty}^{\infty} dx(x - x_0)^2|\psi(x)|^2}{\int_{-\infty}^{\infty} dx|\psi(x)|^2}, \\ \sigma_{\xi_0}^2 &= \frac{\int_0^{\infty} d\xi(\xi - \xi_0)^2|\hat{\psi}(\xi)|^2}{\int_0^{\infty} d\xi|\hat{\psi}(\xi)|^2}. \end{aligned} \tag{6}$$

For any wavelet $\psi_{m,n}$ of the family, the standard deviations are given as $\sigma_x = a_0^m \sigma_{x_0}$, and $\sigma_\xi = \sigma_{\xi_0}/a_0^m$. Although the size of support of the wavelet[2] determines the full extent of the function's

[2]The support of f is defined as the closure of the set $\{x : f(x) \neq 0\}$.

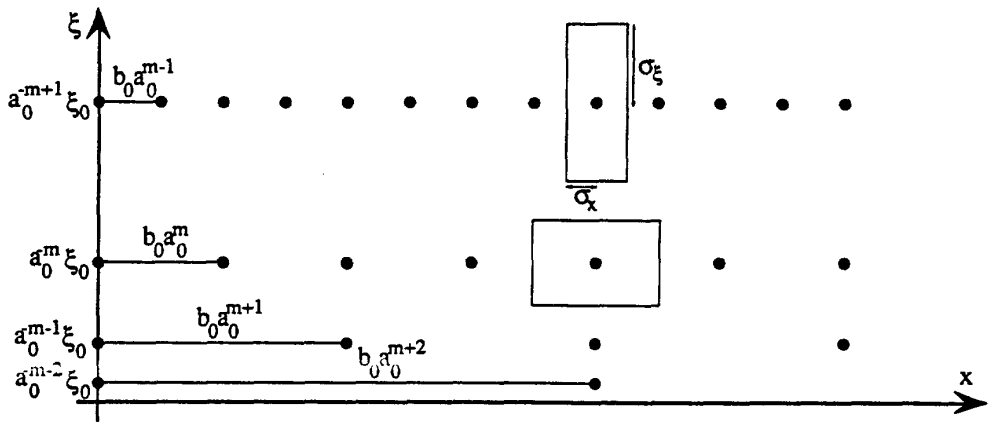

Figure 1: The discrete wavelet family determines the hyperbolic lattice of the phase space, corresponding to the points where wavelets are the most concentrated (Adapted from [Daubechies, 1990]).

window which contributes to the wavelet coefficient, the most important contribution comes from the points inside the width of the wavelet. We define the width of the wavelet as $2\sigma_x$ in the time domain and $2\sigma_\xi$ in the frequency domain. Hence, the size of the cell where the wavelet is concentrated is given by

$$[<x> -\sigma_x, <x> +\sigma_x] \times [<\xi> -\sigma_\xi, <\xi> +\sigma_\xi]. \tag{7}$$

If we identify a_0^{-m} with a constant multiple of the frequency, then it follows from (7) that the time–side of the window narrows at high frequencies and widens at low frequencies. This is the zoom–in and zoom–out capability of the wavelet transform. The areas of all cells are the same, and they cannot be arbitrarily small.

Intuitively speaking $\langle f, \psi_{m,n} \rangle$ then represents the "information content" in f near time $a_0^m nb_0$ and near the frequencies $\pm a_0^{-m}\xi_0$ (Fig. 1). If f itself is "essentially localized" on two rectangles in the time–frequency space, meaning that, for some $0 < \Omega_0 < \Omega_1 < \infty, 0 < T < \infty$

$$\int_{\Omega_0 \le |\xi| \le \Omega_1} d\xi |\hat{f}(\xi)|^2 \ge (1-\delta) \| f \|^2$$

$$\int_{|x| \le T} dx |f(x)|^2 \ge (1-\delta) \| f \|^2,$$

where δ is some small number, then this intuitive picture suggests that only those $\langle f, \psi_{m,n} \rangle$ corresponding to m, n for which $(a_0^{-m}nb_0, \pm a_0^{-m}\xi_0)$ lies within or close to the box $[-T,T] \times ([-\Omega_1, -\Omega_0] \cup [\Omega_0, \Omega_1])$ are needed to reconstruct f to a very good approximation. A theorem by Daubechies [2] states that this is indeed true, thereby justifying our intuitive picture (see Appendix).

3 Object Localization in Frequency Domain

In this section we will show that objects are localized in space and frequency domains, and accordingly the above theorem can be applied. It is important to keep in mind that we need the

representation which will contain information for recognition, rather than reconstruction purposes. Accordingly, we have different requirements than if the representation were intended for reconstruction. There are two assumptions which are essential at this point:

Assumption 1. Visual objects are localized in the space domain. This simply means that at some distance, the image of the entire object will fit within the retina. As the system moves farther and farther away, the object spans smaller and smaller portions on the retina.

Assumption 2. The recognition system is robust, i.e. it can tolerate small deformations of the image of an object. As the object becomes larger and larger, the allowed deformations become larger and larger.

Let us consider the consequences of the first of above two assumptions. We assume an one–dimensional retina, long enough so that transients due to the boundaries do not affect the results of the wavelet transform, i.e. the filter works in the steady state. Let us also assume that we have an isolated object with compact support of $T > 0$ points on an empty background. We want to calculate the wavelet transform of this object. For the time–limited function, the energy of the wavelet coefficients is spread over all frequencies. This means that we will have non-zero wavelet coefficients for any size of the wavelet support. However, not all of them contain the information about the object, as it can be seen by the following analysis. The analysis is specific for the case of discrete functions and a discrete wavelet transform.

In Eq. (2), if we have increase the scale parameter m, the wavelet spreads out in time. The calculated wavelet coefficient from Eq. (3) takes into account only long–time behavior of f. Changing the summation variable in (3), we can see that dilating the wavelet corresponds to contracting the function f. For discrete functions, contracting implies decreasing the resolution, and since the object has a compact support, it will finally shrink to one point. At this coarsest scale we will have convolution of the wavelet with just one point of the object, which is equivalent to multiplying by a constant. This scale is obtained when the width of the wavelet is equal to the width of the object and is equal to

$$m_T = \log\left(\frac{T}{2\sigma x_0}\right) / \log(a_0). \tag{8}$$

Therefore, this scale (or related frequency) is the coarsest scale (lowest frequency) in the phase space that contains information about the object and the scales (frequencies) beyond it are useless for this object. The size of object's support determines a lower bound $\Omega_0 > 0$ on the object frequencies.

Since we do not want the background to interfere with the information about the object, even this frequency is too low. It was noticed earlier that each wavelet coefficient carries the same amount of information (time–bandwidth product) about the entire function. However, if we consider the coefficients which characterize the particular time–limited object, the boundary coefficients carry less information about the object because they are corrupted by the background. Since the ratio of corrupted coefficients relative to the total number of object's coefficients increases with scale, it is better to set the lower bound of frequencies higher than $\Omega_0 > 0$. This bound should be in the point where the error introduced by background overrides the error introduced by abandonin low–frequency band.

Each object, depending on its size, has proper range of scales which contain the information about it. The finest scale is determined by the sampling of the original signal and the coarsest by 8. It is important to notice that the energy is not equally distributed on all of these scales. The

energy of the function is defined as $\sum_k |f(k)|^2 = \sum_\kappa |\hat{f}(\kappa)|^2$. Due to the time scaling property of the Fourier transform ($f(t/a) \leftrightarrow a\hat{f}(a\xi)$, $a > 0$), as an object becomes larger and larger, more and more of its energy will be located in the low frequency region.

In order to enable the recognition, we have to allow some amount of variation for all objects. Due to lighting conditions, occlusions, damages, parallaxes, etc., there is a very low probability that the object will appear in the image the same ever again.

Let us say that two objects (on an empty background) are represented by $f_1, f_2 \in \ell^2(\mathbb{Z})$. We will say that the two objects are similar if the ℓ^2 norm[3]

$$\| f_1 - f_2 \| = \left(\sum_{k=-\infty}^{\infty} |f_1(k) - f_2(k)|^2 \right)^{1/2} = \epsilon_T < \infty.$$

This means that we can add a function f_3 with norm $\| f_3 \| = \epsilon_T$ to the function representing the object, and still be able to recognize the object. Since the energy of the function is the square of its ℓ^2 norm, the function f_3 has a small energy and it is the same both in the time and the frequency domains. Since the amount of energy in the high frequencies for a large object is relatively small, the high frequency part of the spectrum and the corresponding wavelets are the most susceptible to the addition or subtraction of energy. Hence, this part of the spectrum of the original object may be safely excluded from the object's analysis. This conclusion appears reasonable because we do not expect the shape of a large object to depend significantly on its small details.

Thus, the quantity ϵ_T determines the width of the narrowest wavelets used in the wavelet analysis of the object. The associated frequency $\Omega_1 > 0$ is the upper bound on the object frequencies. Obviously, $\Omega_0 < \Omega_1$. The conclusion is that for the purpose of recognition every object *can be localized in the frequency domain* as well.

Abandoning the high frequencies of the object's spectrum produces a generalization, and therefore, recognition now becomes *classification*, i.e. we are now dealing with determining the object's class, not a particular instance of it[4]. Object classification requires localization both in the spatial and frequency domains.

Narrow objects span small portions of the retina, but require more dense calculations of wavelet coefficients because their phase space cell falls in the region of wavelets with small support (see Figure 2). The reverse is true for wide objects: They span large portions, but require less dense calculation of wavelet coefficients. Henceforth, the system may use similar amount of input data over all scales for the classification purposes. For the reasons of convenience, we assume that all object cells in the phase space have the same number of coherent states.

Since it is not known in advance where on the retina an object will appear, the inputs to the classifiers, i.e. the cells of phase space, should be taken all over the retina. They are overlapping and their centers are positioned on a discrete lattice. The amount of spacing of this lattice is determined as follows. The size of an object determines the absolute amount of the acceptable deformation. However, this determines the amount of spacing between the centers of input arrays for different classifiers as well. Namely, the deformation includes the part of the object which may be missing due to the discrete spacing of the classifiers.

The conclusion is that the visual recognition system for each object is justified to process just M points instead of all N points of the retina ($N \gg M$). However, the number of cells in the phase space is $N - M$. Instead of a huge amount of input points to a single classifier, we obtained a huge

[3]We are interested only in ℓ^2 norm similarity measure because the wavelet transform is performed on entire functions. Other types of similarity measures would require figure–ground segregation and further manipulation of particular objects independently of one another.

[4]Notice that the requirement for robustness by definition implies classification.

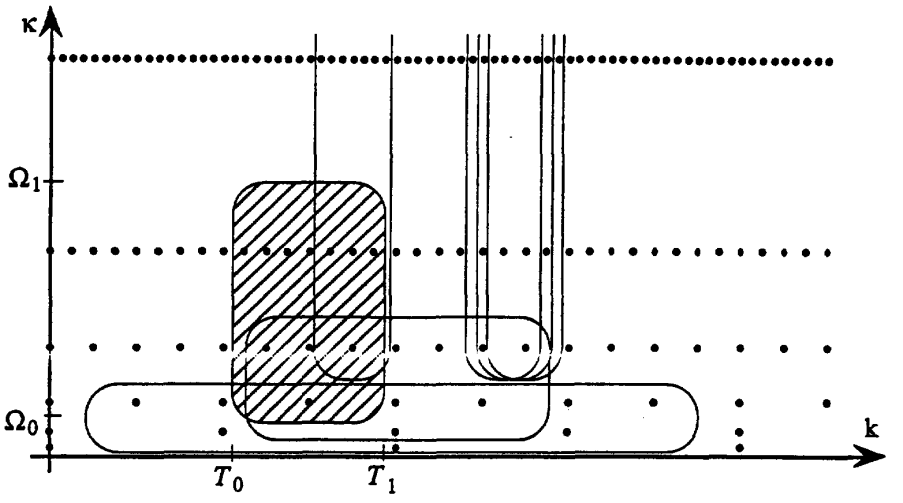

Figure 2: Depending on the object's location and size, the above analysis says which are the coordinates where the object is located in phase space. Everything outside is irrelevant under the given assumptions. Regardless of the object's size, all cells include the same number M of wavelet coefficients at $\log_2 M$ different scales.

amount of cells which will be one by one fed to one classifier or each cell may have one classifier associated with it. We can notice that it is not necessary at each portion of the retina to process cells at all scales. Finer scales may be processed only in the central portions of the retina, and the rest may be accounted by the eye movements.

Until this point, we had no special requirements on the wavelet family. The goal was to decompose the input function into different frequency or scale subbands, apply subsampling and feed these wavelet coefficients into subsequent processing modules. However, if the object moves into the neighboring cell of phase space (at the same scale), it will obviously produce different wavelet coefficients, due to discrete sampling [5]. Therefore, the wavelet transform with sampling does not possess the property of *translation invariance*. We need sampling because that reduces amount of data kept about the objects at any scale. We can assume that each classifier works independently of others, and then the problem with translation variance would not exist as long as the wavelet transform produces unique coefficients for each object. However, the problem arises with directing the attention of subsequent modules to particular cells of phase space, as it will be seen in the next section.

4 Analyzing Objects in Wavelet Phase Space

The purpose of the first processing stage is to separate object descriptions at a proper spatial location and the range of scales or frequencies. Each object class is represented as a collection of wavelet coefficients in the phase space, determined by the object's size. How are these phase space cells further processed?

There is no need for combining these descriptions because they have been combined in the input image and purposefully separated as cells of the phase space. Their corresponding classifiers

do not need information from other space–frequency locations in order to work properly. In fact, combining them will obstruct the process of classification. Hence, each cell is processed separately.

How do subsequent modules know which cell of phase space contains the interesting objects? —This information may be obtained either indirectly from other sources or directly from the cells themselves. We are interested in the second case. Based on some criteria, the cells have to compete for attention of the subsequent modules.

Each object in one–dimensional image is defined by its left and right edge, which should be of oposite sign. The slope of these edges gives an estimate of the height of their object, and this is the first saliency criterion: *At any particular scale, the most salient objects are those which are the highest.* If there are several objects of the similar height at the same location but different scale, then *the most salient object is the one at the coarsest scale.*

Within any phase space cell, the most important edges and associated wavelet coefficients are those which are spatially the most distant and therefore peripheral. The close ones are the remanents of the objects from finer scales. Hence, the inner edges must be neglected for the purpose of finding the most salient objects at a particular scale and across the scales.

We are still justified in keep the same number of coefficients for object at any scale, since this is determined by Nyquist sampling frequency. However, it is important to shift the sampling grid so that the maximum information about the edges is obtained.

5 Discussion of Further Work

We have shown that each object in the input image belongs to the different cell in phase space. This simplifies the process of learning and recognition. In addition, this approach determines the unique method of figure–ground segregation.

If the biological vision system will utilize this approach to the problem of complexity of visual tasks, we cannot expect such exact regularity. Spatial frequency channels probably will not be in exactly prescribed subbands; they may adaptively move within the non–sampled phase space. The bandwidth, though, should remain approximately the same on the logarithmic scale.

Particular elements of the object class can be distinguished only by the difference in the higher frequency wavelets, which were abandoned when creating the class. Since wavelets are spatially localized too, this implies that particular objects differ from one another by some sub-object. Associating wavelet collections for the different objects/sub-objects creates the representations of particular objects. Thus, another issue that requires further work is introducing (spatial) order in the set of icons to represent patterns of objects.

One of the interesting issues is what happens if the largest wavelet coefficients of the object fall further away in the frequencies than it is predetermined by spatial size of the object. Then it must be that the object consist of many small objects. The examples may be textures and Marroquin's pattern.

The purpose of this work is to conduct the quantitative analysis and to develop computer programs for practical testing of the approach. We are primarily interested in developing a suitable framework for object recognition, and thus further work will be along the lines outlined in [6].

Appendix

(Theorem 3.5.1 in [2])

THEOREM (Daubechies) *Suppose that the $\psi_{m,n}$ constitute a frame with frame bounds A,B, and suppose that*

$$|\psi(x)| \leq C(1+x^2)^{-\alpha/2}, \quad |\hat{\psi}(\xi)| \leq C|\xi|^{\beta}(1+\xi^2)^{-(\beta+\gamma)/2},$$

with $\alpha > 1, \beta > 0, \gamma > 1$. Then, for any $\epsilon > 0$, there exists a finite set $\mathcal{B}_\epsilon(\Omega_0, \Omega_1; T) \subset \mathbb{Z}^2$ so that, for all $f \in L^2(\mathbb{R})$,

$$\| f - \sum_{(m,n)\in\mathcal{B}_\epsilon(\Omega_0,\Omega_1;T)} \langle f, \psi_{m,n}\rangle \widetilde{\psi_{m,n}} \| \leq \sqrt{\frac{B}{A}} \left[\left(\int_{\substack{|\xi|<\Omega_0 \\ \text{or } |\xi|>\Omega_1}} d\xi |\hat{f}(\xi)|^2 \right)^{1/2} + \left(\int_{|x|>T} dx |f(x)|^2 \right)^{1/2} + \epsilon \| f \| \right].$$

The theorem says that, given a time–limited and band–limited function f, the wavelet decomposition does not spread it around in the phase space, i.e. keeps it localized. It guarantees the "graceful degradation" of the results of wavelet transform: By removing the complete scale of wavelet coefficients (one row), or coefficients of all scales at any position (one column), the introduced error will be proportional to the amount of energy in that row or column.

References

[1] DAUBECHIES, I., "The Wavelet Transform, Time–Frequency Localization and Signal Analysis," *IEEE Trans. Inform. Theory*, Vol.36, No.5, pp.961–1005, September 1990.

[2] DAUBECHIES, I., *Ten Lectures on Wavelets*, CBMS, SIAM Publ., to appear.

[3] DEVALOIS, R.L., and K.K. DEVALOIS, *Spatial Vision*, Oxford University Press, 1988.

[4] JUDD, J.S., *Neural Network Design and the Complexity of Learning*, The MIT Press, Cambridge, MA, 1990.

[5] MALLAT, S., "Zero–Crossings of a Wavelet Transform," *IEEE Trans. Inform. Theory*, Vol.37, No.4, pp.1019–1033, July 1991.

[6] MARSIC, I., and E. MICHELI–TZANAKOU, "A Framework for Object Representation and Recognition," *Proc. Int. Joint Conf. Neural Networks*, Baltimore, MD, June 1992.

[7] TSOTSOS, J.K., "Analyzing Vision at the Complexity Level," *Behavioral Brain Sci.*,Vol.13, pp.423–469, 1990.

[8] WITKIN, A.P., "Scale Space Filtering: A New Approach to Multi–Scale Description," in *Image Understanding 1984*, S. Ullman and W. Richards (eds.), Ablex Publ. Co., Norwood, NJ, pp.79–95, 1984.