Report Number: WUCS-90-23

1990-12-01

A Multi-Scale Approach for Recognizing Complex Annotations in Engineering Documents.

Andrew Francis Laine, William Ball, and Arun Kumar

This paper describes a novel method of character recognition targeted for extracting complex annotations found in engineering documents. The results of this work will make it possible to capture the information contained in documents used to support facilities management and manufacturing. The recognition problem is made difficult in part because characters and text may be expressed in arbitrary fonts and orientations. Our approach includes a novel incremental strategy based on the multi-scale representation of wavelet decompositions. Our approach is motivated by biological mechanisms of the human visual systems. Using wavelets as a set of basis functions, we may decompose... **Read complete abstract on page 2.**

A Multi-Scale Approach for Recognizing Complex Annotations in Engineering Documents.

Andrew Francis Laine, William Ball, and Arun Kumar

Complete Abstract:

This paper describes a novel method of character recognition targeted for extracting complex annotations found in engineering documents. The results of this work will make it possible to capture the information contained in documents used to support facilities management and manufacturing. The recognition problem is made difficult in part because characters and text may be expressed in arbitrary fonts and orientations. Our approach includes a novel incremental strategy based on the multi-scale representation of wavelet decompositions. Our approach is motivated by biological mechanisms of the human visual systems. Using wavelets as a set of basis functions, we may decompose an image into multiresolution hierarchy of localized information at different spatial frequencies. Wavelet bases are more attractive than traditional hierarchical bases because they are orthonormal, linear, continuous, and continuously invertible. The multi-scale representation of wavelet transforms provides a mathematically coherent basis for multi-grid techniques. In contrast to previous ad-hoc approaches, our method promises a practical solution embedded in a unified mathematical theory. A feasibility study is described in which several hundred characters extracted from engineering drawings were recognized without error by a neural network trained using multi-scale representations from a class of 36 distinct alphanumeric patterns. We observed a 16-fold reduction in the amount of information needed to represent each character for recognition. These results suggest that high reliability is possible at a reduced cost of representation.

A MULTI-SCALE APPROACH FOR RECOGNIZING COMPLEX ANNOTATIONS IN ENGINEERING DOCUMENTS

Andrew Laine, William Ball and Arun Kumar

WUCS-90-23

December 1990

Department of Computer Science
Washington University
Campus Box 1045
One Brookings Drive
Saint Louis, MO 63130-4899

Submitted to IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Submitted as a REGULAR paper to

IEEE Computer Society Conference on
Computer Vision and Pattern Recognition

A MULTI-SCALE APPROACH FOR RECOGNIZING COMPLEX ANNOTATIONS IN ENGINEERING DOCUMENTS

Andrew Laine

Computer and Information Sciences Department
Computer Science and Engineering Building, Room 301
University of Florida
Gainesville, FL 32611-2024

William Ball and Arun Kumar

Department of Computer Science
Washington University
Campus Box 1045, Bryan 509
Saint Louis, Missouri 63130-4899

ABSTRACT

This paper describes a novel method of character recognition targeted for extracting complex annotations found in engineering documents. The results of this work will make it possible to capture the information contained in documents used to support facilities management and manufacturing. The recognition problem is made difficult in part because characters and text may be expressed in arbitrary fonts and orientations. Our approach includes a novel incremental strategy based on the multi-scale representation of wavelet decompositions.

Our approach is motivated by biological mechanisms of the human visual system. Using wavelets as a set of basis functions, we may decompose an image into a multiresolution hierarchy of localized information at different spatial frequencies. Wavelet bases are more attractive than traditional hierarchical bases because they are orthonormal, linear, continuous, and continuously invertible. The multi-scale representation of wavelet transforms provides a mathematically coherent basis for multi-grid techniques. In contrast to previous ad-hoc approaches, our method promises a practical solution embedded in a unified mathematical theory.

A feasibility study is described in which several hundred characters extracted from engineering drawings were recognized without error by a neural network trained using multi-scale representations from a class of 36 distinct alphanumeric patterns. We observed a 16-fold reduction in the amount of information needed to represent each character for recognition. These results suggest that high reliability is possible at a reduced cost of representation.

1. Introduction

This paper describes a novel method of pattern recognition targeted for recognizing complex annotations found in paper documents. Our investigation is motivated by the problem of automating the interpretation of maps and engineering drawings. The results of this work will make possible the capture of information contained in documents supporting facilities management and manufacturing.

Fundamental to achieving an autonomous production capability is the development of a reliable method for recognizing the characters and symbols that are contained within a drawing. While recent methods of character recognition [1,2,11] have been successful in reading printed text from books, extracting annotations within the context of engineering drawings and maps requires a more general and robust method. In particular, the problems of orientation (recognizing text placed non-horizontally) and feature extraction (separation of text from graphics) remain unsolved.

We intend that the recognition capability described in this paper provide input to a system capable of "understanding" identified annotations through higher level reasoning. Such a system would process input annotations in the context of a specific engineering domain and generate an electronic description for both graphical and textual information contained within an engineering drawing. Some previous contributions to engineering drawing interpretation are summarized in a survey paper by Nagendra and Gutar [12]. In addition recent works by Hishihara and Ikeda [13], Lysak and Kasturi [14], and Whitaker and Huhns [15] describe efforts towards the development of an autonomous interpretation system for maps and engineering drawings.

We present a novel method of character recognition that we believe will make such a technology feasible and attractive (low cost). Our method includes an incremental strategy for recognizing characters based on the multi-scale representation of wavelet decompositions [3-6]. Using wavelets as a set of basis functions, we may decompose an image into a mul-

tiresolution hierarchy of localized information at different spatial frequencies. Wavelet bases are more attractive than traditional hierarchical bases because they are orthonormal, linear, continuous, and continuously invertible. The multi-scale representation of wavelet transforms provides a mathematically coherent basis for multi-grid techniques. In contrast to previous ad-hoc approaches, our method promises a practical solution embedded in a unified mathematical theory.

We have developed a novel incremental strategy that utilizes the mathematical continuity between hierarchical levels of wavelet decompositions. Similar to traditional coarse to fine matching strategies, we first attempt to recognize coarse features within low frequency levels of the wavelet transform. If higher resolution is required to resolve an ambiguity, we add incrementally to the representation, the finer features of a pattern available at higher frequency levels. Choosing wavelets (or analyzing functions) that are simultaneously localized in both space and frequency, results in a powerful methodology for image analysis. The inner-product of a signal x with a wavelet f ($< x, f > = (2\pi)^{-1} < \hat{x}, \hat{f} >$) reflects the character of x within the time-frequency region where f is localized ($\hat{f}$ and $\hat{x}$ are the Fourier transforms of the analyzing function f and the signal x). If f is spatially localized, then 2-D features such as shape remain preserved in the transform space! Our approach is motivated in part by recently discovered biological mechanisms of the human visual system [9,10]. Both multiorientation and multiresolution are features of the human visual system.

Our strategy is to disregard information within the high frequency levels of the basis, and achieve recognition using only time-frequency information concentrated within the low frequency bands. Thus, we represent a pattern at the hierarchical level corresponding to the lowest frequency band possible, such that the fundamental shape of each character is not lost. The wavelet decomposition allows us to discard redundant transform coefficients by decimation, between descending levels of the hierarchy. This results in a 4 fold difference (reduction) in the number of transform coefficients stored within each level of the hierarchical basis. By accomplishing recognition using information available at the lower levels of the

hierarchy, we require fewer transform coefficients to represent each character. This is desirable from an information theoretic point of view in that the technique converges towards a minimal form of representation without compromising resolution.

We present experimental results, in which several hundred characters extracted from real engineering drawings were recognized without error by a neural network trained with dilated wavelet representations (space-frequency coefficients) from a class of 36 distinct alphanumeric patterns. Our investigation shows that recognition using dilated representations of a space-frequency transform, not only yielded high reliability, but resulted in a 16 fold reduction in the amount of information (number of bits) needed to accomplish our recognition task.

In the next section we describe the problem and motivate our approach. In Section 3 we present a detailed overview of our recognition strategy. Next, Section 4 presents a mathematical formulation of a multi-scale hierarchical basis and shows the construction of the analyzing functions used in our investigation. Finally, Section 5 presents a summary and discussion of our results.

2. Motivation and Problem Statement

Manufacturing companies such as McDonnell Douglas and utility companies such as Southwestern Bell Telephone have large quantities of engineering drawings and facilities maps that exist exclusively as paper documents. Such companies have an urgent need to convert these paper documents into electronic form. For example, McDonnell Douglas will be required to supply existing and future engineering drawing pertaining to the F-15 fighter in an electronic form, by the year 1995. Over half of the facilities maps describing the Southwestern Bell Telephone network remain as paper documents. Southwestern Bell Telephone estimates that converting the remaining maps into electronic form will cost at least 58 million dollars, and require ten years, using manual and existing semi-automated methods.

We propose to develop a production quality image understanding system that will convert such paper-based engineering drawings and facilities maps into an electronic form, that may be stored, retrived and updated via conventional and geographical databases oriented for CAD/CAM and facilities transactions respectively. The development of such a technology would not only benefit manufacturing and utility companies (mentioned above), but may also be used to check the design of chemical plants and to model circuit diagrams for diagnosis.

Figure 1(a) shows the topology of telephones poles, conduits, and manholes of a urban neighborhood in Houston. Figure 1(b) contains the codings for the cables and connections boxes that overlay the network topology shown in Figure 1(a). The extraction of both textual and graphical features are necessary to "understand" such facility maps. Given the state of the art, finding a reliable method of character recognition remains fundamental to accomplishing an autonomous production capability. We present a novel method of character recognition that we believe will make such a technology feasible and attractive (low cost). In the next paragraph we identify four fundamental problems of character recognition that are addressed by the methodology and techniques described in this paper.

Recognizing characters and symbols in the context of drawings and maps, requires that four sub-problems be solved: (1) Font invariance, (2) Intensity invariance, (3) Scale invariance, and (4) Orientation invariance. Our method of recognition must be invariant to font style because a drawing may have been updated by several engineers over it's lifetime. Note that the facilities map shown in Figure 1(a) is over 30 years old! Invariance to intensity and size is needed since several writing instruments (eg. typewriters, rapidograph pens, pencils) may have been used to draw and maintain (update/delete) the original drawing. Invariance to orientation is needed to handle symbols and text placed at non-horizontal positions: The engineer/draftsman, may have placed text and symbols in non-horizontal or non-vertical orientations to keep information localized while making the best use of available "white space" within a drawing. In the next section, we discuss our approach towards these four goals.

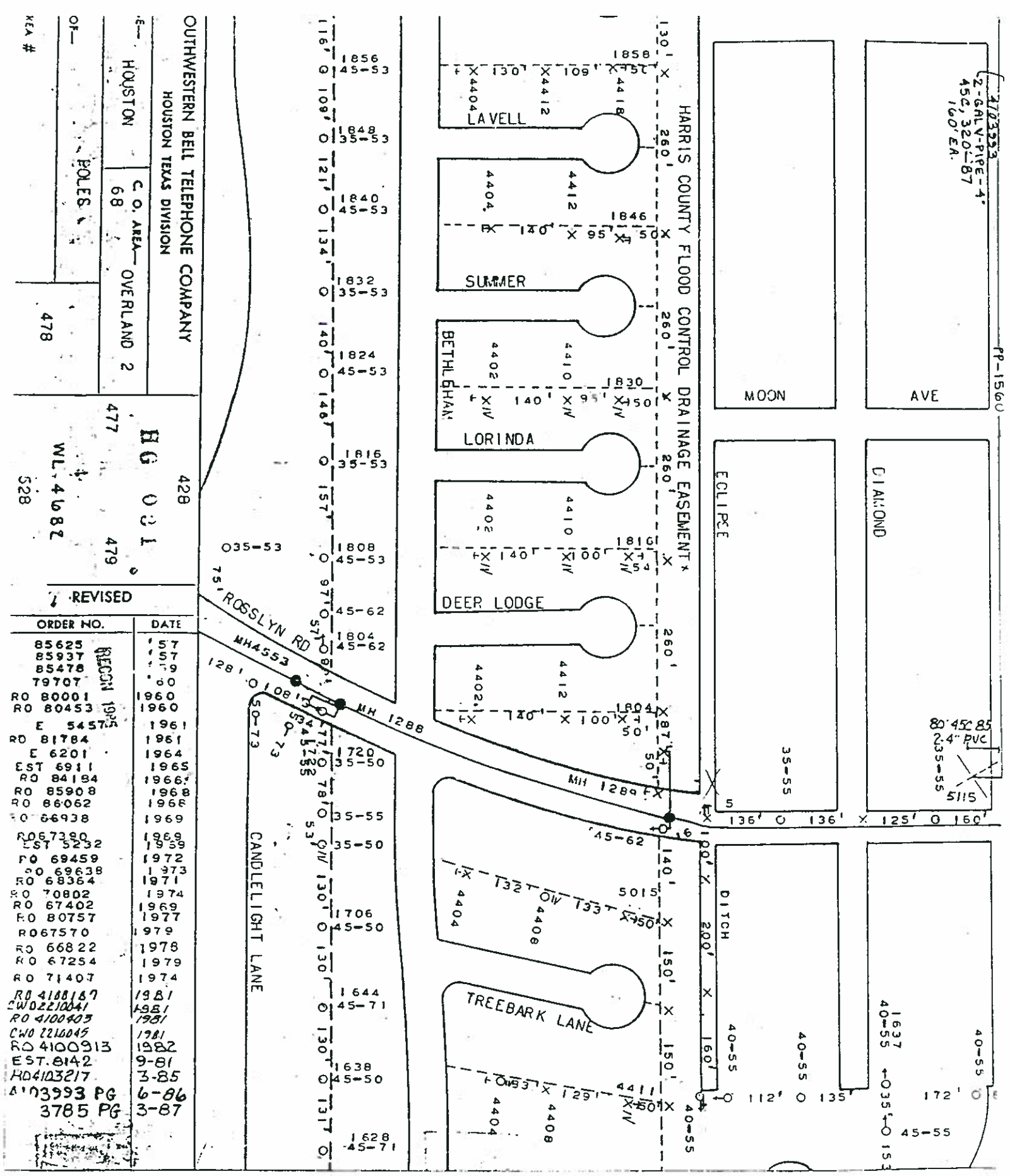


Figure 1(a). Network topology for an urban neighborhood.

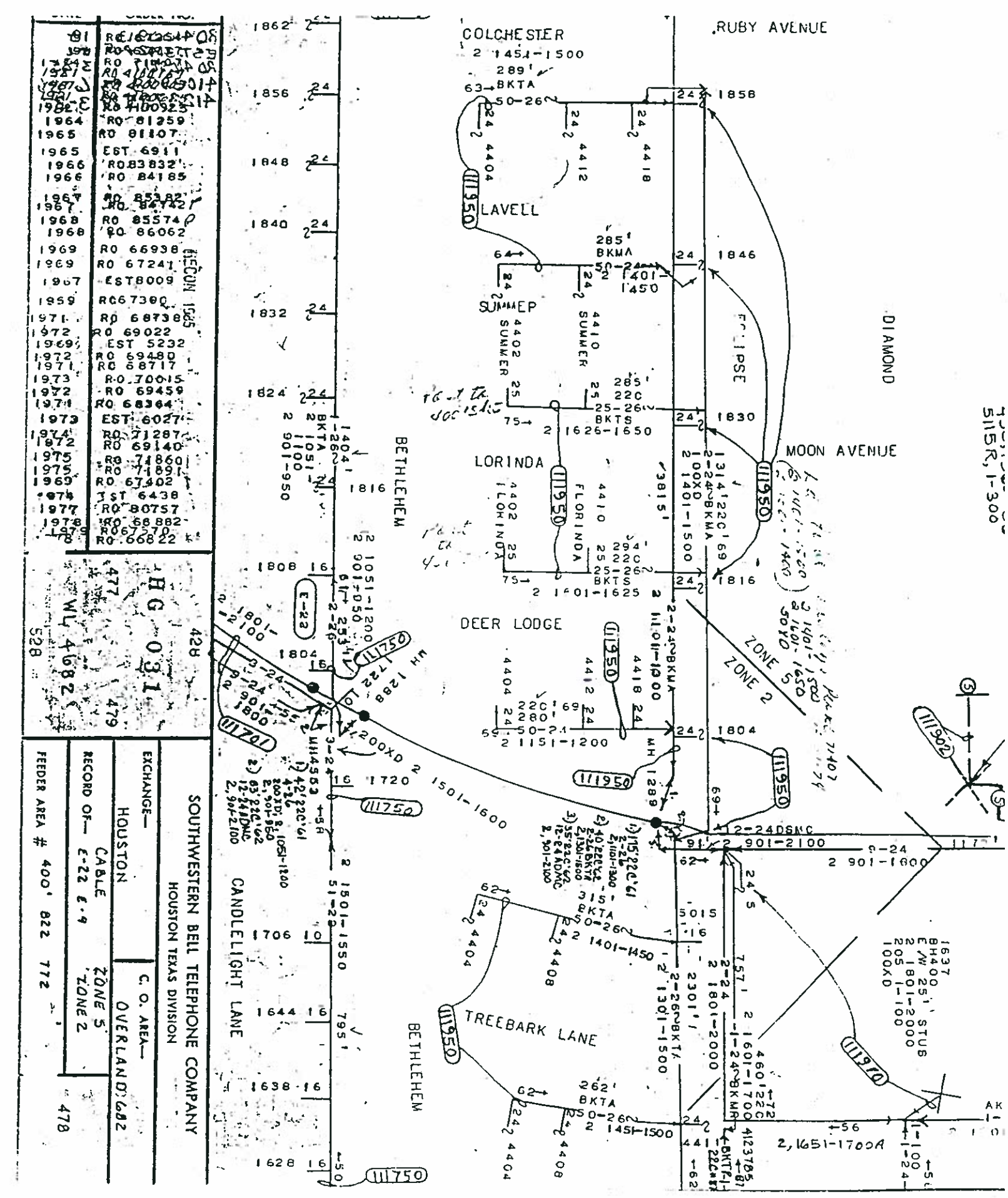


Figure 1(b). Codings for connections and wiring on network topology.

3. Strategy and System Overview

This section presents a strategy for character recognition exploiting a multiresolution hierarchical basis. First, we present a general overview the character recognition system accomplished to date. A more formal description of the hierarchical basis used in our investigation is presented in Section 4.

We have obtained a number of facility management maps from Southwestern Bell Telephone Company to use as test cases during our investigation of automatic image understanding of maps and drawings. We utilize both software and hardware to allow the maps to be scanned into the computer, including image processing routines that may need to be applied to the image before the "understanding" process begins. We have initiated the development of a database of characters extracted from these (real) maps and drawings. This database forms the foundation upon which the accuracy of our character recognition process is evaluated. We have constructed a neural network that is being trained to recognize characters using the existing database of extracted characters. The neural network has been trained to recognize characters using a minimal form of representation obtained by a method of space-frequency analysis, previously applied to the compression of digital pictures.

In the large, our research is separated into two phases. The first phase is concerned with low level problems of feature extraction. The second phase accomplishes high level reasoning by connecting extracted "words" and symbols with meaningful entities that may be encapsulated within a database. Unfortunately, the efforts and results of the second phase are beyond the scope of this discussion. In the paragraphs below, we present an overview of our method for reliable feature extraction and character recognition.

We now describe a method to extract characters from a paper-based engineering drawing and how they are transformed into a "compacted" representation that is used to train a neural network for general classification. Our approach, summarized in Figure 2., consists of six steps:

SYSTEM OVERVIEW

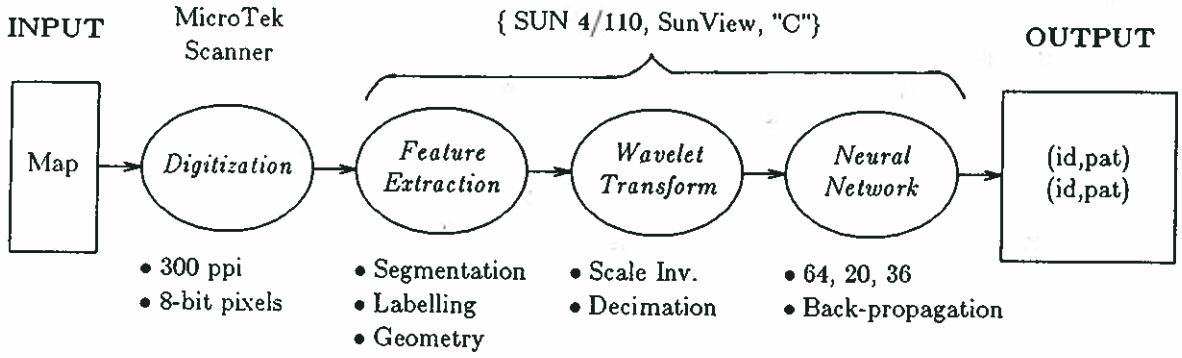


Figure 2. A Testbed for Character Recognition.

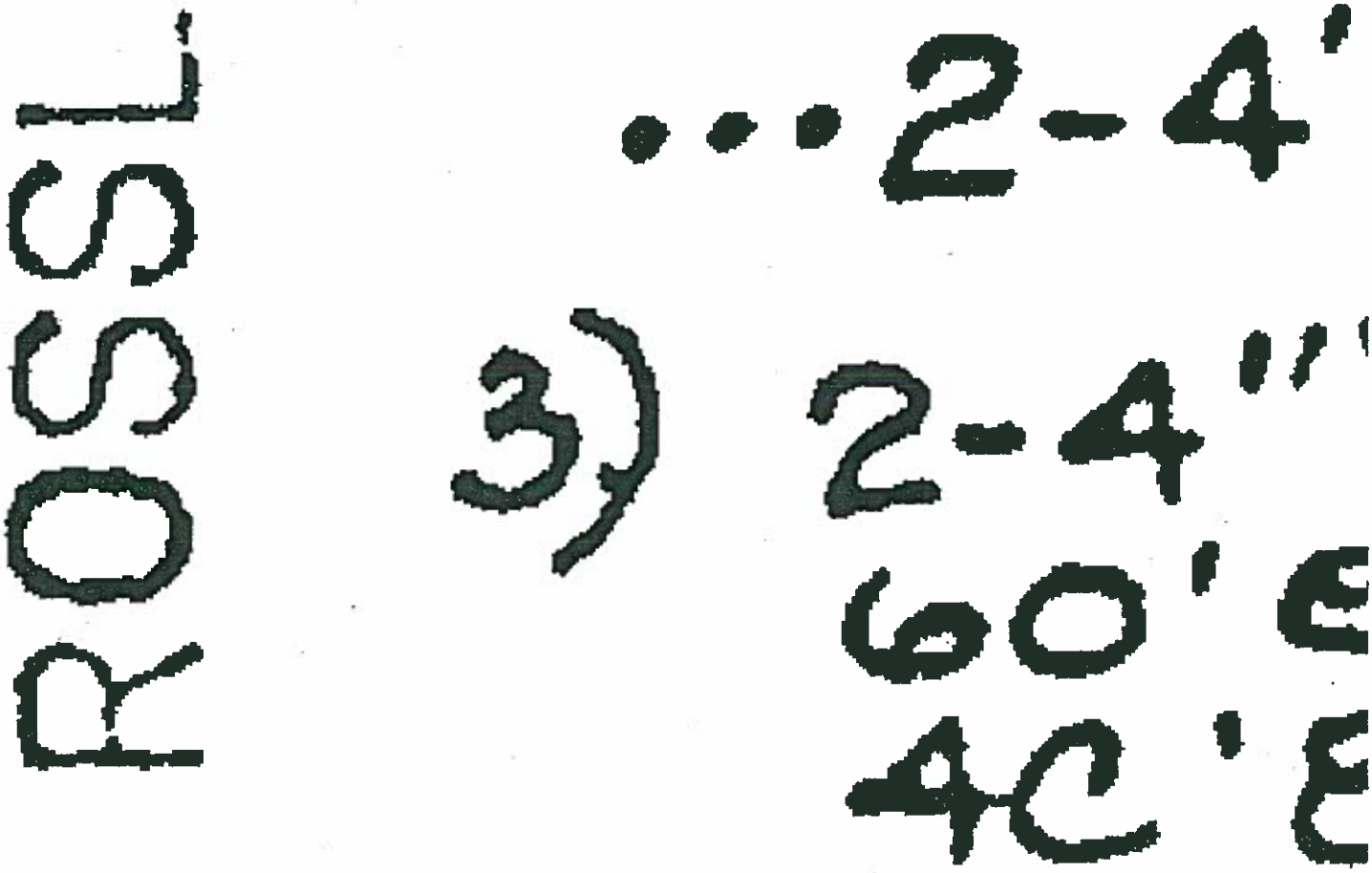


Figure 3. Sample digitization (300 ppi) of characters extracted from a facility map.

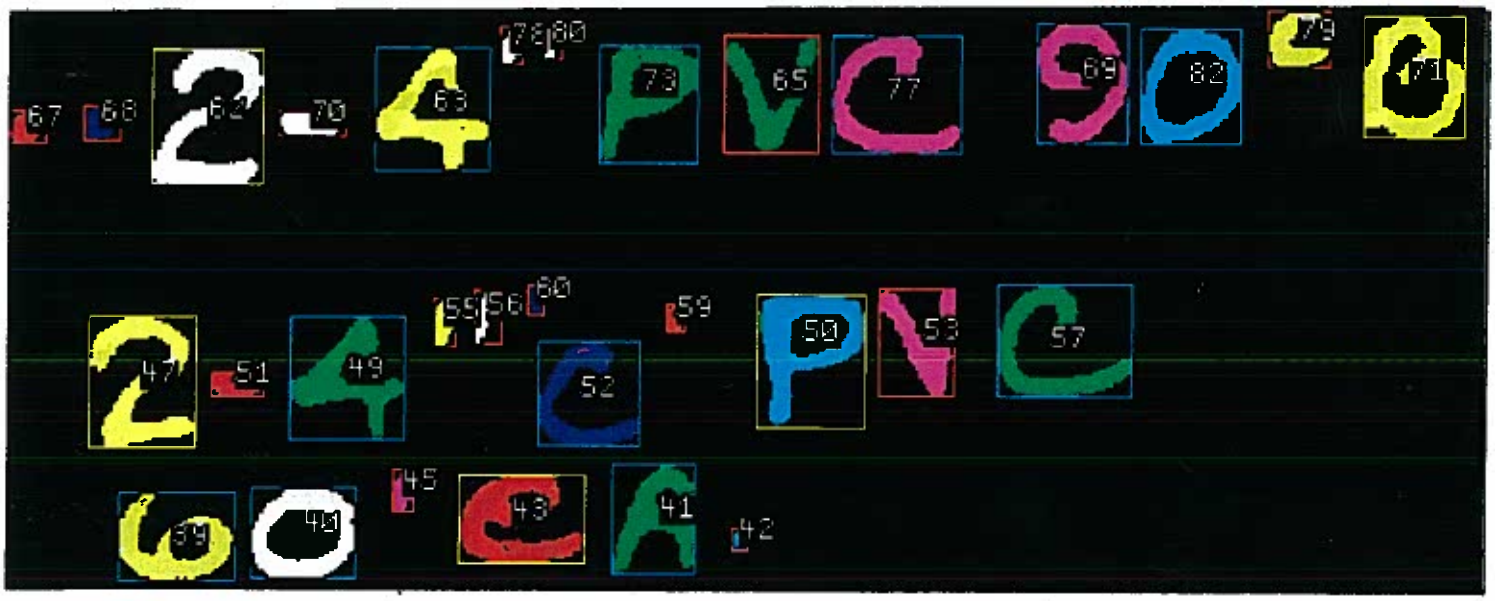


Figure 4. Segmented and labelled components delineated by minimum bounding rectangles.

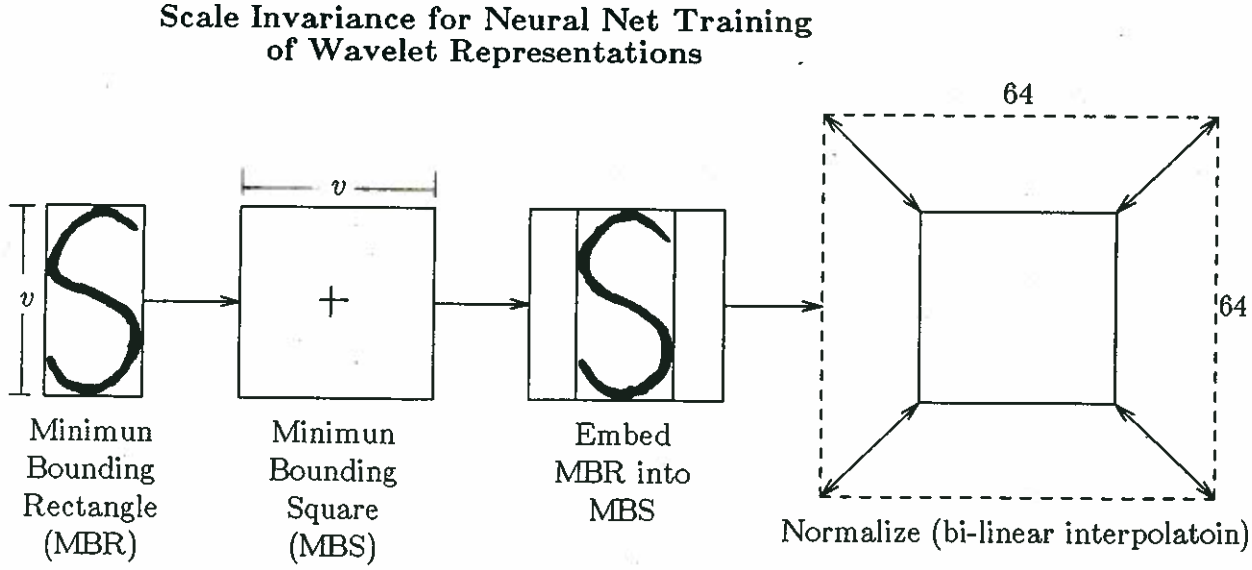


Figure 5. A shape preserving technique for scale invariance.

(1) ***Digitize original drawing.*** First, the map is digitized at 8-bits per pixel (256 grey levels) at a resolution of 300 pixels per inch, using a Micro-Tek 300Z scanner. Figure 3 shows some digitized characters sampled from one of our maps.

(2) ***Segment and label connected components.*** Next, segmentation is accomplished by labelling 8-connected components having similar grey-level intensity with a unique identifier. The user may select a range of grey-level intensity to group similar pixels, for each segmentation. As a preprocessing step, we may engage traditional image processing techniques to "clean-up" noisy drawings.

(3) ***Compute geometric properties for each labelled segment.*** In this step, we compute for all segments a set of geometric properties including area, centroid, maximum height and width. These properties will be used later on to separate segments of characters from graphics, establish baselines and for clustering local groups of characters and symbols into meaningful "chunks" of information. In Figure 4, each segment is shown labelled and enclosed within a minimum bounding rectangle. Figure A.1 shows the complete set of geometric properties computed for each segment.

(4) ***Apply geometric constraints to classify segments.*** Next, we use a logical combination of the geometric properties computed from the previous step to classify each segment into one of three disjoint partitions: characters, noise or graphics. Very small segments are most likely noise (or punctuation), very large segments are most likely pieces of graphics. Experimentally we have observed that segments having specific height to width ratio and normalized area may be coarsely classified as characters, and marked for further processing.

(5) ***Transform character segments into a multiscale representation.*** Each segment marked as a character is decomposed by a hierarchical basis into a decimated representation. In our study, we used a multiresolution decomposition that is closely related to the wavelet decomposition called the the Psi-transform [5]. In contrast to wavelet basis, the basis (or analyzing functions) used in the Psi-decomposition are non-orthogonal. Both methods of time-

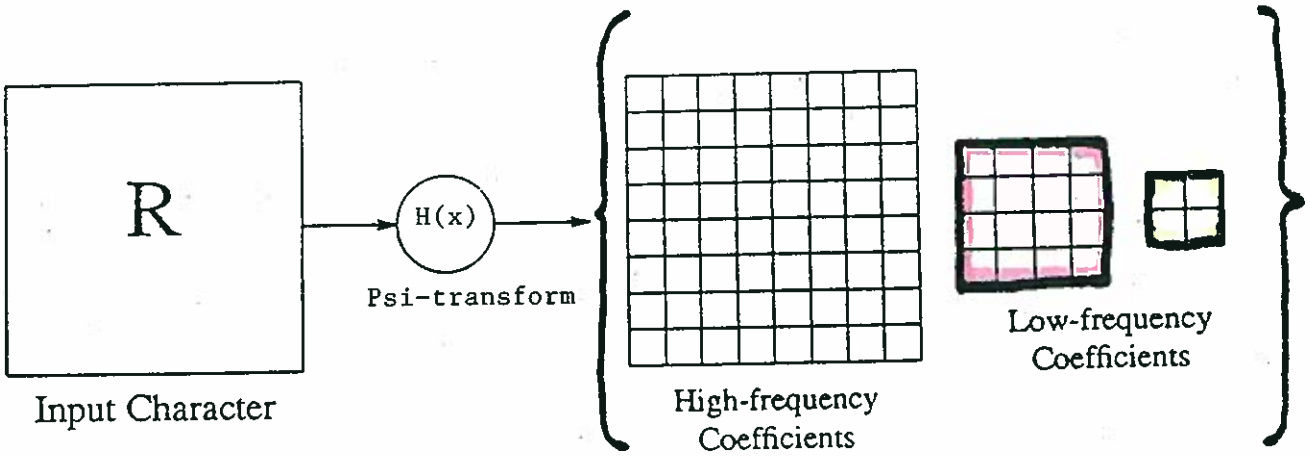
frequency analysis have been previously used to decompose and quantize digital signals for compression.

Before transformation, each character segment is normalized in scale to fit a minimum bounding square of 64 pixels. This is accomplished by first constructing a minimum bounding rectangle (MBR) and identifying the longest edge, parameter v, as shown in Figure 5. A minimum bounding square (MBS) is then allocated to match the length of the longest edge of the MBR. Next, the MBR is embedded within the MBS. (Note that in fitting the MBR into the MBS, at most one degree of freedom will exist due to the geometric constraint set by the parameter v.) Once embedded, a character may be shrunk or enlarged without distorting it's original shape by the method of bilinear interpolation, as shown in Figure 5.

The Psi-transform decomposes the information within the bounding square into a multi-scale space-frequency representation. Our strategy is to disregard information within the high frequency levels of the basis, and achieve recognition using only time-frequency information concentrated within the lower frequency bands. We represent each pattern at the hierarchical level corresponding to the lowest frequency band such that the fundamental shape of each character is not lost. The Psi-decomposition allows us to discard redundant transform coefficients by decimation, between descending levels of the hierarchy. This results in a 4 fold difference (reduction) in the number of transform coefficients stored within each level of the hierarchical basis. By accomplishing recognition using information available at the lower levels of the hierarchy, we require fewer transform coefficients to represent each character. This is desirable from an information theoretic point of view in that the technique converges towards a minimal form of representation without compromising resolution.

A SPACE-FREQUENCY TRANSFORM

FOR CHARACTER RECOGNITION



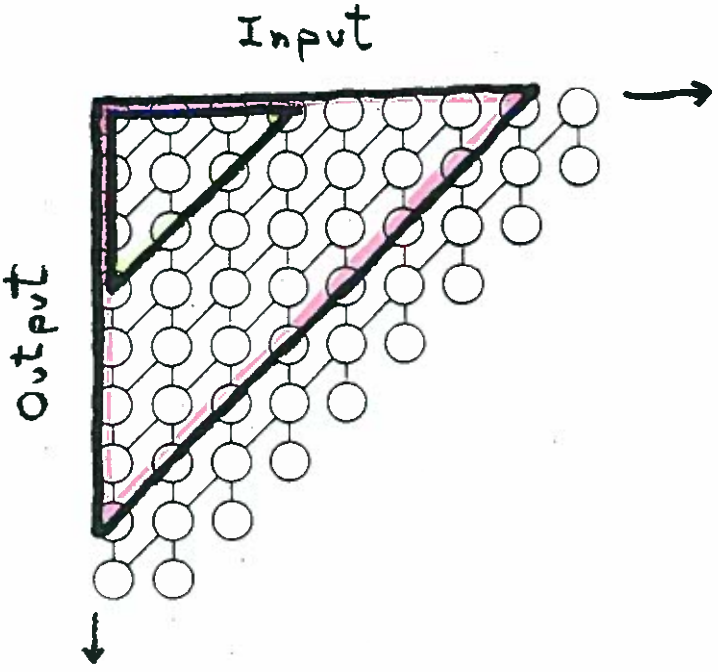


Figure 6 An incremental strategy for multi-scale character recognition.

In our investigation, we were able to reduce the number of transform coefficients needed to discriminate a class of 32 distinct alphanumeric patterns by a factor of 64! (4096 original vs. 64 decimated). Figure 11 shows a set of transform coefficients for a sample character, computed from the set of non-orthogonal basis functions shown in Figures 7 and Figure 8. Note that dilations ϕ_{-1}, ϕ_{-2}, and ϕ_{-3} are subsampled by a factor of 2 in each dimension.

Another advantage of using a non-orthogonal basis for multi-scale decomposition is that the shape (geometry) of a signal remains localized in transform space. In contrast to the Fourier transform, both frequency content *and* the temporal (spatial) evolution of a (non-stationary) signal are captured by the Psi-decomposition. The Psi-transform is linear, continuous, and continuously invertible. Thus, if we find that more information is needed to discriminate a character or symbol, we may incrementally add higher frequency information to lower frequency representations until it is resolved. Figure 6 shows how hierarchical multi-scale representations of such transforms might map onto the topology of a systolic neural network to support an incremental strategy.

(6) ***Exploit existing neural network pattern recognition technology.*** Next, we use the low frequency components of the Psi-transform to train a neural network to recognize the 36 alphanumeric characters [A-Z,0-9]. We use the method of back-propagation for training. Since we have reduced the representation needed for each character and symbol, our neural network needs only 64 input units (one for each transform-coefficient), rather than 4096 input units, that would have been needed to train a network using the original representation. The rest of the two-layer network consists of 20 hidden units and 36 output units. Figure A.2 shows a sample of the input file format used to train our neural network.

Scaling, translation, and multiple fonts are handled quite well by the procedures described above. However, there remains one difficulty: How do we handle rotated characters? When annotations are made, engineers may choose write in whatever empty white

space is available, no matter what the orientation. The steps described in the previous section assume that characters are mostly horizontal in orientation.

Our approach to handling non-horizontal characters is unique. We first identify clusters of characters (the "words" used to describe features for a specific engineering domain) by searching the drawing from left to right starting at the upper leftmost corner. Next, based on the aggregate direction of each cluster, we rotate each character within a cluster the degrees necessary to re-orient the character along a horizontal line. This approach has the advantage in that it is less expensive computationally and more reliable than methods that have attempted to train neural networks to recognize each character in it's many possible orientations. We have employed techniques developed by Kasturi [7] to help identify graphics using the precomputed spatial and geometric properties of labelled segments.

To complete the overview of our approach, we briefly describe the strategy for the second phase of our research. In phase two, we address the higher level task of putting the characters together to form sets of meaningful annotations for the graphics within the drawing. Baseline identification, word boundaries, character/word ambiguities, and domain vocabulary are all considered here. Thus, the second phase of our investigation focuses on processing the recognized symbols and characters into meaningful pieces of information for a given engineering domain. The key objective is the development of automated methods that identify and resolve any problems of ambiguity that may arise when processing complex descriptions or notations. To gain insight into this problem, we have consulted with domain experts (engineers and users) from industries committed to CAD/CAM, facilities engineering, chemical engineering and circuit design engineering. Thus, the second phase will integrate specific domain knowledge obtained from such experts into a high level reasoning system. For example, a non-monotonic reasoning system may be used to analyze the information gathered from a drawing and provide this information (in symbolic and/or geographical form) to a database.

In the next section, we present a mathematical formulation of the hierarchical basis and analyzing functions used to accomplish the method of mult-scale character recognition described above.

4. A Hierarchical Basis for Character Recognition

We applied a multiresolution decomposition using non-orthogonal basis called the Psi-transform. The Psi-transform, proposed by Frazier and Jawerth [5], is a simple yet rigorous method for time-frequency analysis of nonstationary signals. Psi-transforms are a family of transforms strongly related to the family of wavelet transforms. In both transforms we write a signal as a weighted sum of certain "elementary" functions [6]. However, wavelet decompositions require analyzing functions (bases) that are orthonormal.

In applying the Psi-transform, a signal is written as a weighted sum of certain "elementary" functions. Similar to wavelet and Gabor methods of time-frequency analysis, we look at a signal through a set of windows. However, the Psi-transform uses an entire scale of windows of different widths. Wider windows are used to capture slowly varying features, while narrow windows track sharper details. The set of windows is obtained from a single parent function (sometimes called the "mother wavelet") through a process of dilation and translation.

Let $\mathbf{Z}, \mathbf{R}$ and $\mathbf{C}$ denote the sets of integers, reals and complex numbers. For sake of clarity, we restrict our discussion to the case of one-dimensional signals. However, the methods and formulations generalize to $\mathbf{C}^n$ function spaces. We define $\hat{\phi}_\nu(\omega) \equiv \hat{\phi}(2^{-2}\omega)$, $\nu \in \mathbf{R}$, $\omega \in \mathbf{R}^n$, as a dilation of ϕ in the frequency-domain, and $\phi_\nu(t) \equiv 2^{n\nu}\phi(2^\nu t)$, $\nu \in \mathbf{R}$, $t \in \mathbf{R}^n$, as a dilation of ϕ in the time-domain (where "^" and "v" denote the forward and inverse Fourier transforms, $\hat{f}(\omega) = \langle f(t), e^{j\omega t}\rangle$, and $f(t) = (2\pi)^{-1}\langle \hat{f}(\omega), e^{-j\omega t}\rangle$). In addition let us define, $\phi_{\nu k}(t) \equiv 2^{n\nu/2}\phi(2^\nu t - k)$, as a dilation-and-translation of ϕ in the time domain. Thus, the symbols ν and k denote dilation and translation parameters respectively.

A signal f may be expressed as a sum of *synthesizing functions* from the set $\mathbf{S} = \{\Psi_{mk}, \psi_{vk}\}_{v,k}$, $m, v \in \mathbf{Z}$, where m is a fixed but arbitrary integer, $v > m$, and $k \in \mathbf{Z}^n$. The functions Φ_{mk} in S are the translates (in $\mathbf{R}^n$) of the single "parent" function Φ_{m0}, while ψ_{vk} are translated-and-dilated versions of a single function ψ.

The Psi-transform of a signal f is the countably-infinite sequence $\mathbf{T}f = (\langle f, \Phi_{mk}\rangle, \langle f, \phi_{mk}\rangle)_{v,k}$ of inner-products of f with *analyzing functions* from the set $\mathbf{A} = \{\Phi_{mk}, \phi_{vk}\}$. All functions in $\mathbf{S} \cup \mathbf{A}$ are defined from $\mathbf{R}^n$ to C. If S and A are appropriately chosen, then any signal or function $f : \mathbf{R}^n \to \mathbf{C}$ in $L^2(\mathbf{R})$, can be written as follows:

$$f(t) = \sum_{k \in \mathbf{Z}} \langle f, \Phi_{mk}\rangle \Psi_{mk}(t) + \sum_{v=m}^{\infty} \sum_{k \in \mathbf{Z}} \langle f, \phi_{vk}\rangle \psi_{vk}(t). \qquad (1)$$

The weighted sum of inner-products in (1) is called the *Psi-decomposition* of the signal f. The symbol k denotes a vector $(k_1, \cdots, k_n) \in \mathbf{Z}^n$. Similarly, arguments to the functions f, Φ_{mk}, Ψ_{mk}, ϕ_{vk}, and ψ_{vk}, are vectors in $\mathbf{R}^n$.

The expression in (1) is not unlike the decomposition of the Fourier transform of a signal f. Since $\hat{f}(\omega) = \langle f, e^{j\omega t}\rangle$, we can write:

$$f(t) = \int_{\mathbf{R}^n} \langle f(t), e^{j\omega t}\rangle \frac{e^{j\omega'}}{(2\pi)^n} d\omega. \qquad (2)$$

As such, we have formulated the function f as a weighted sum of certain "elementary" functions - the complex exponentials. The complex exponentials are perfectly localized in the frequency domain, while they range widely in the time domain. The weights $\langle f(t), e^{j\omega t}\rangle$ in (2) are the Fourier coefficients of f, and are related to the projections of the signal f onto the complex exponentials $e^{j\omega t}$.

Similarly, the Psi-decomposition (1) of f, may be formulated as a weighted sum of the synthesizing functions in S. Where the weights in (1) constitute the countable sequence $\mathbf{T}f = (\langle f, \Phi_{mk}\rangle, \langle f, \phi_{vk}\rangle)$ of complex coefficients. The numbers in the sequence $\mathbf{T}f$ are related to the projections of f upon the vectors in the set A comprising the analyzing functions. The complex numbers $\langle f, \phi_{vk}\rangle = (2\pi)^{-n} \langle \hat{f}, \hat{\phi}_{vk}\rangle$ depend upon the behavior of f only in

those regions of the time and frequency domains where the functions ϕ_{vk} and $\hat{\phi}_{vk}$ are appreciably non-zero. If the functions in **A** are chosen such that they are simultaneously localized in both time and frequency, then $\mathbf{T}f$ will yield a time-frequency representation of f [4].

This decomposition provides a solid mathematical found on which to embed multiresolution techniques. In contrast to traditional hierarchical bases, the Psi-decomposition defines a *linear, continuous, and invertible* transformation. In current wavelet research, much time is spent in "choosing" analyzing functions most appropriate for a particular application. In the next section, we describe the analyzing functions we constructed as a set of basis for our hierarchical decomposition. We will see that the first term of (1) results from a lowpass filter operation, while the second term results from a series of bandpass filter operations.

4.1. Generation of Analyzing Functions

As mentioned above, the analyzing and synthesizing functions (sets **A** and **S**) can be generated by several methods [16] [4]. In our study, we used cosine-of-log functions for both analyzing and synthesizing functions, because they were relatively simple to design. In the paragraphs below, we present the mathematical formulation for the construction of the basis set.

Let us cover the frequency line with functions from the set $W = \{\hat{\Theta}_m\} \cup \{\hat{\theta}_v\}$ of *windows*, as shown in Figure 7a, so that:

$$\hat{\Theta}_m + \sum_{v=m+1}^{\infty} \hat{\theta}_v \equiv 1, \ \forall \omega \in \mathbf{R}. \tag{3}$$

By the *support* of a function $f : \mathbf{R} \rightarrow \mathbf{C}$ we mean the topological closure of the set $\{x \in \mathbf{R} \mid f(x) \neq 0\}$. No function may have bounded support in both time and frequency domains, as a consequence of the time-frequency uncertainty principle [17]. However, it is possible to design an analyzing function (ϕ) that has bounded support in the frequency domain, and *localized* support in the time domain, as shown below.

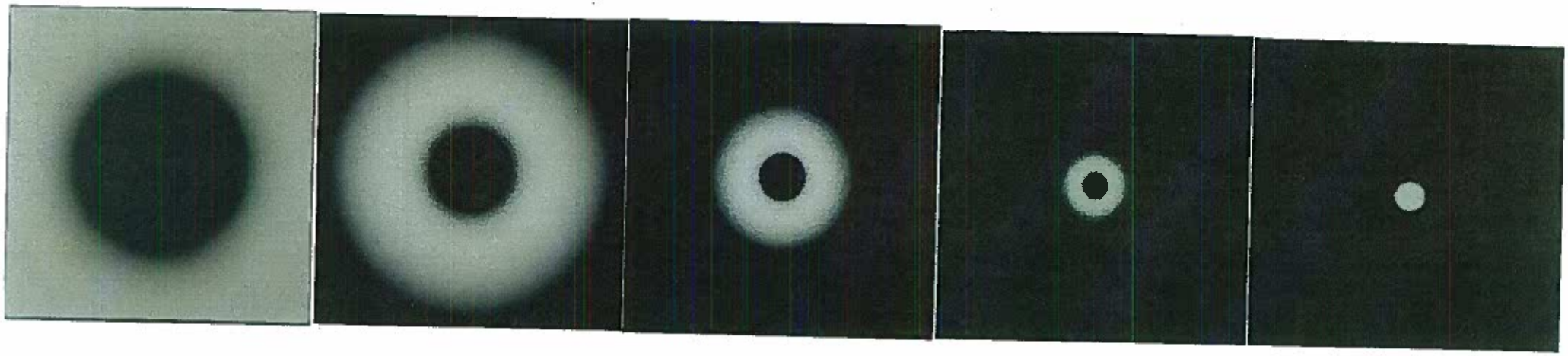


Figure 7b. The Set of Analyzing Functions in the Frequency Domain, $\hat{\phi}_{+1}$, $\hat{\phi}_0$, $\hat{\phi}_{-1}$, $\hat{\phi}_{-2}$, $\hat{\Phi}_{-3}$.

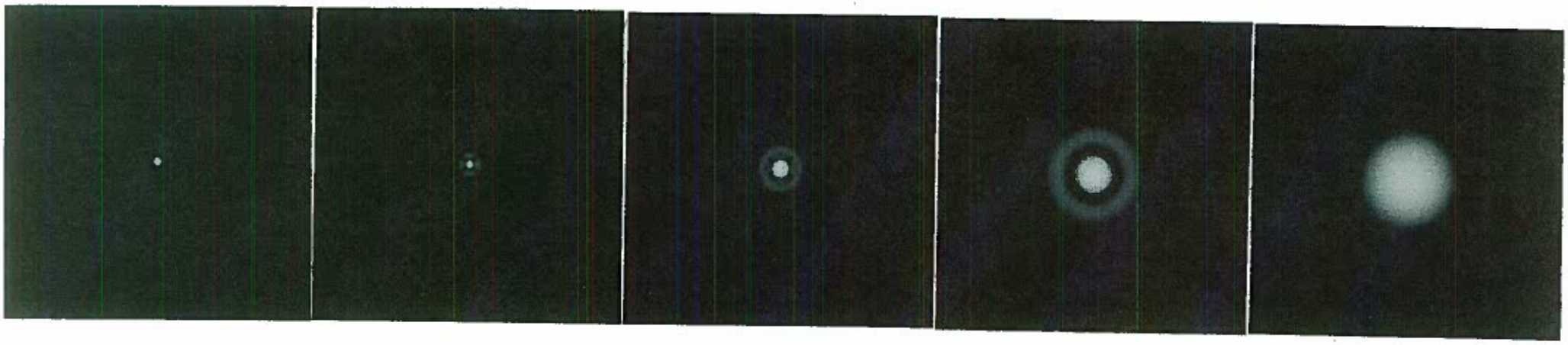


Figure 8. The Set of Analyzing Functions in the Time Domain, ϕ_{+1}, ϕ_0, ϕ_{-1}, ϕ_{-2}, Φ_{-3}.

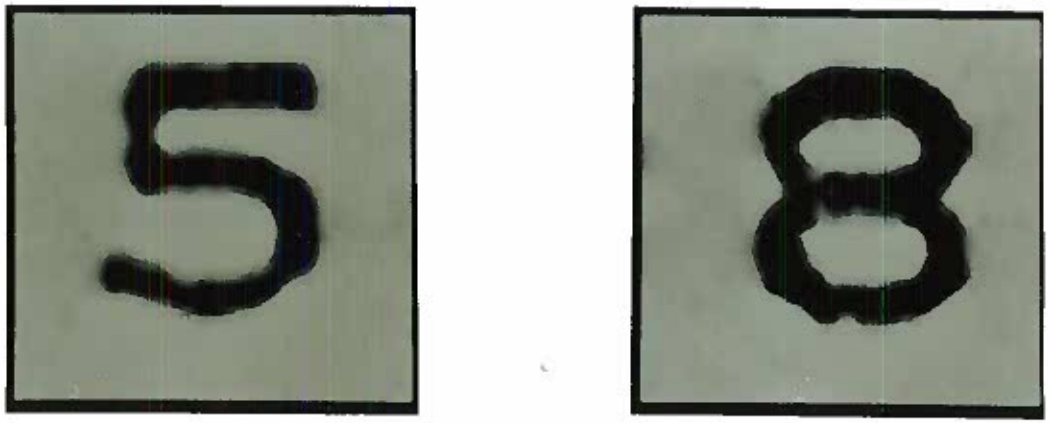


Figure 9. Sample Input Characters Extracted from Engineering Drawings.

In equation (3) we assume that the *support* $\hat{\Theta} \subseteq \{ \| \omega \| \leq \pi \}$, and *support* $\hat{\theta} \subseteq \{ \frac{\pi}{4} \leq \| \omega \| \leq \pi \}$. It follows trivially that

$$f = f\hat{\Theta}_m + \sum_{v=m+1}^{\infty} f\hat{\theta}_v, \quad \forall \omega \in \mathbf{R}. \tag{4}$$

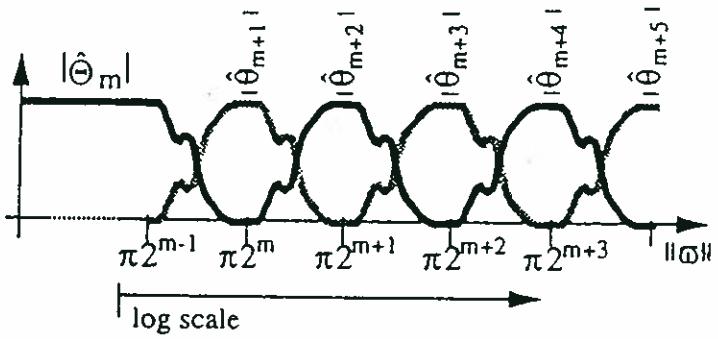


Figure 7a A set of windows covering the frequency line

In this investigation, we used a simple construction where the "window-function" $\hat{\theta}$ was chosen to be a raised cosine pulse whose argument is the logarithm of the frequency variable,

$$\hat{\theta}(\omega) \equiv \begin{cases} \frac{1}{2}\left(1 - \cos(\pi \log_2 \| \omega \|)\right) & , \frac{\pi}{4} \leq \| \omega \| \leq \pi \\ 0 & , \text{otherwise} \end{cases} \tag{5}$$

Thus the sum of all the dilations of $\hat{\theta}$ is exactly 1.0, for all $\omega \neq 0$. $\hat{\Theta}$ was chosen to satisfy (3), over the frequency interval $[-\pi, \pi]$. The functions $\hat{\phi}$ and $\hat{\psi}$ were chosen to be the square-root of $\hat{\theta}$. The functions $\hat{\Theta}$ and $\hat{\theta}$ were factored such that the DC cap functions $\hat{\Phi} = \hat{\Psi}$ and $\hat{\phi} = \hat{\psi}$. Thus, in our study the analyzing functions and synthesizing functions were the same.

The sets containing the analyzing and synthesizing functions, S=A, are constructed by first choosing an integer m, and then computing the translates $\{ \Phi_{mk} : k \in \mathbf{Z} \}$. (The integer m determines the number of hierarchical levels created by dilations of the decomposition. Experimentally, we fixed $m = -3$, generating three dilated (subsampled) levels within a five level hierarchical decomposition, as shown in Figure 11.) Finally, we compute the translates and dilates $\{ \phi_{vk} : v = m+1, m+2, \ldots; \text{and } k \in \mathbf{Z} \}$.

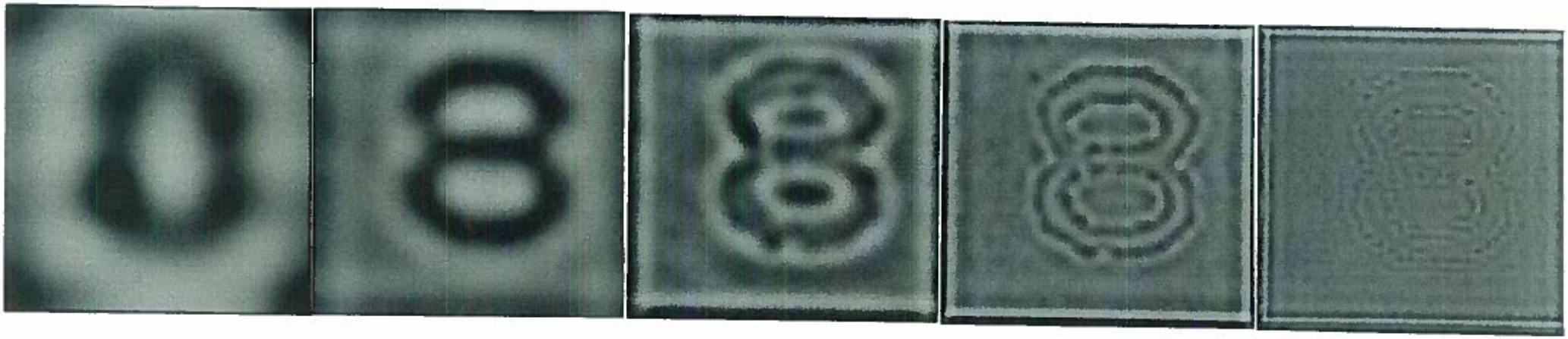


Figure 10. Transform Coeffecients, $(\hat{f}\hat{\Phi}_{-3})^{\vee}$, $(\hat{f}\hat{\phi}_{-2})^{\vee}$, $(\hat{f}\hat{\phi}_{-1})^{\vee}$, $(\hat{f}\hat{\phi}_{0})^{\vee}$, $(\hat{f}\hat{\phi}_{+1})^{\vee}$.

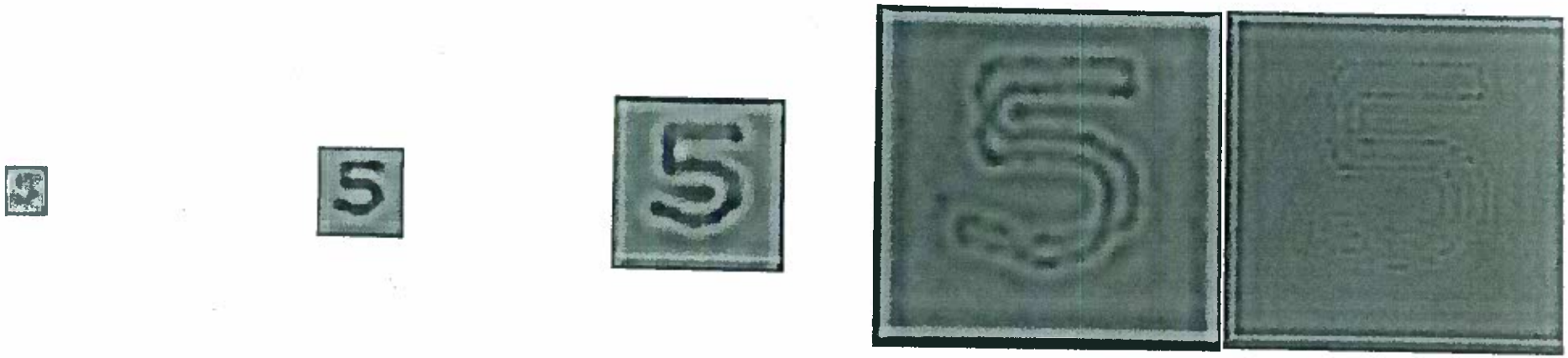


Figure 11. Transform Coeffecients, $(\hat{f}\hat{\Phi}_{-3})^{\vee}$, $(\hat{f}\hat{\phi}_{-2})^{\vee}$, $(\hat{f}\hat{\phi}_{-1})^{\vee}$, $(\hat{f}\hat{\phi}_{0})^{\vee}$, $(\hat{f}\hat{\phi}_{+1})^{\vee}$.
Shown without decimation.

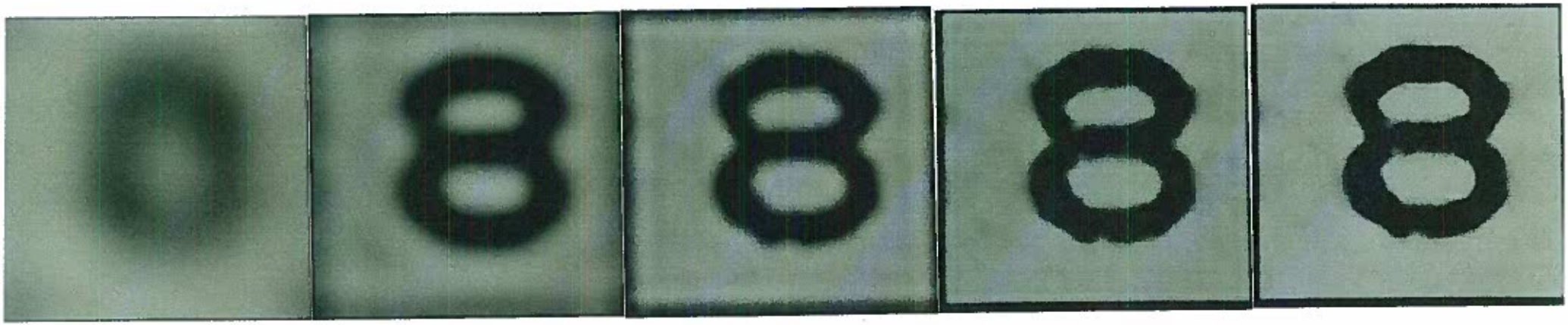


Figure 12. Incremental Reconstruction, from low to high frequency bands, $\hat{f}\hat{\Phi}^2_{-3} + \sum_{\nu=-2}^{1} \hat{f}\hat{\phi}^2_{\nu} = \hat{f}$.

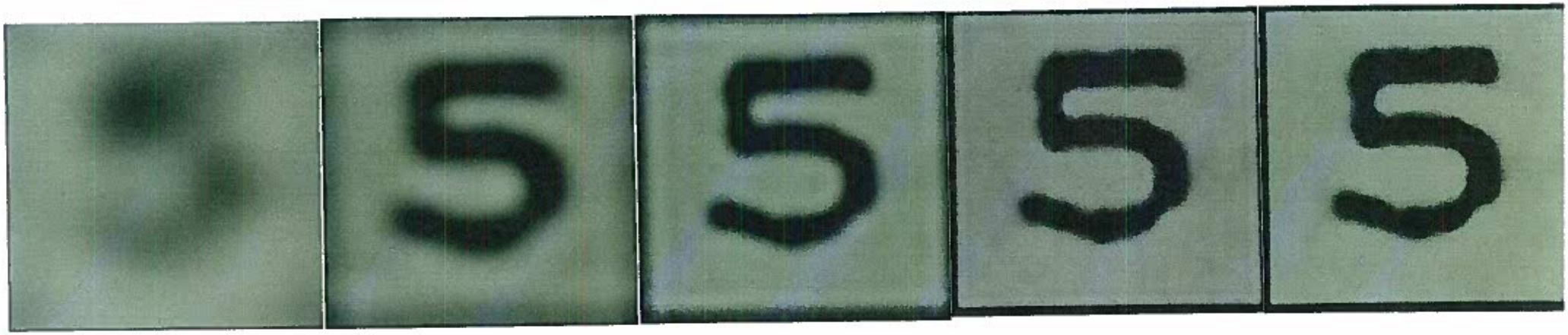


Figure 13. Incremental Reconstruction, from low to high frequency bands, $\hat{f}\hat{\Phi}^2_{-3} + \sum_{\nu=-2}^{1} \hat{f}\hat{\phi}^2_{\nu} = \hat{f}$.

In figures 7 and 8 we show the functions $\Phi_m, \phi_{m+1}, \phi_{m+2}, \ldots, \phi_1$, for $m = -3$ in the frequency and spatial domains respectively. Since $\hat{\Phi}$ and $\hat{\phi}$ are discontinuous at the boundaries of their support, the functions Φ and ϕ are localized in space, but not perfectly concentrated. Figure 9 shows two sample characters ("8" and "5") extracted from engineering drawings and normalized in scale by the method described in Section 3 to 64 pixels on a side. Figures 10 shows the magnitude of the transform coefficients for each individual level of the hierarchical decomposition. (The coefficient values displayed were biased to a positive scale before computing the magnitude to avoid folding the signal over at zero-crossings.) Note that the number of transform coefficients at frequency level ν is one-forth that at level $\nu+1$. The transform coefficients are all real, since the input image f is real and the analyzing functions (A) are real and symmetric in both frequency and spatial domains, as shown in Figures 7b and 8 respectively. The transform coefficients shown in Figure 11 were computed computed using the same cosine-of-log analyzing functions, but are shown undecimated for exposition. Finally, Figures 12 and 13 shows the incremental reconstruction of the original image patterns shown in Figure 9 from the dilated levels of the transform, proceeding from the lowest space-frequency level Φ_{-3}, to the highest space-frequency level ϕ_1.

In the next section, we discuss the results obtained in our study, and identify areas for further investigation.

5. Summary of Results and Discussion

We have presented a novel method of character recognition based on a mult-scale time-frequency transformation, closely related to the multiresolution wavelet transform. Within each level of the hierarchical decomposition, input patterns were formulated as weighted sums of certain elementary synthesizing functions. Synthesizing functions were constructed from dilated and translated versions of two parent functions, which were shown to be concentrated in both space and frequency domains.

We have begun to build a database of characters, extracted from real maps and drawings obtained from Southwestern Bell Telephone and McDonald Douglas. We have tested over 700 sample characters to date. All characters tested were recognized without error by a neural network trained by backpropagation with low frequency components of the Psi-transform.

Experimentally, we observed a significant reduction in the amount of information needed to represent each character for recognition. We determined that a sixteen fold reduction in the input bandwidth of a neural network, was achieveable by training the network using dilated space-frequency transforms of each of the 36 alphanumeric patterns [A-Z,0-9]. For most characters sampled, we found that representation at dilation level Φ_{-3} (as shown in Figure 10) was sufficient for recognition. Quantitatively, we reduced the number bits required to accomplish recognition from 32,768 bits (original input pattern, 64 x 64 x 8) to 2,048 bits (transform coefficients for dilation Φ_{-3}, 8 x 8 x 32). Thus only 64 input units were needed to configure the network, rather than 4096 which would have been needed to train a network using the original representation. In addition, this greatly reduced the time required to train the network. The network configured as a two-layer 64-20-36, converged within 40 minutes executing on a Sun 4/110 with 16 megabytes of memory.

These results suggest high reliability at a *reduced cost* of representation! While these results are exciting, they are preliminary, and more sample characters need to be collected to rigorously test the method. We estimate that between 2,000 and 4,000 samples need to be collected to rigorously exercise and validate the method. Accordingly, we have developed a semi-automated tool, to improve the reliability of collecting such a large set of sample characters, from a variety of engineering drawings and maps.

If orthogonal analyzing functions can be designed such that they are well localized in both space and frequency, then such a basis (or wavelet) could be applied in a similar fashion, via the wavelet transform. In the near future we plan to use an orthogonal wavelet

basis that is well localized in space. Here, we expect even faster convergence times, since the inputs to the neural network (the transform coefficients) will be orthonormal. In addition we may be able to resolve many more patterns, since each pattern would be trained on an orthonormal basis.

The availability of low-cost high-resolution scanner technology has spurred a great interest in document processing. To realize the potential of this technology, more research will be required to develop new techniques capable of exploiting such high-quality, low-cost digitization. In this context, we claim that the pattern recognition capability and analytic method presented in this paper, can assist not only in the development of a technology for automatic understanding of engineering drawings and maps, but may also be useful for the solution of other problems related to the fields of machine vision and pattern recognition.

Acknowledgements: The authors would like to thank Mr. James Curain of McDonnell Douglas Corporation, Mr. Rudy Reyna of SBC Technology Resources, Inc. and Takuya Kitahara of Mitsubishi Electronics America, Inc. for supporting this research through the Center for Intelligent Computer Systems at Washington University.

Appendix I - References

[1] H.F. Li, R, Jayakumar and M. Youssef, "Parallel Algorithms for Recognizing Handwritten Characters Using Shape Features," *Pattern Recognition*,

[2] David Burr, "Experiments on Neural Net Recognition of Spoken and Written Text," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 36 , No. 7, pp. 1162-1168, July, 1988.

[3] Stephane Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Transactions on PAMI, Vol. 11 No. 7, pp. 674-693, July, 1989.

[4] Ingred Daubechies, "Orthonormal Bases of Compactly Supported Wavelets", Communications on Pure and Applied Mathematics, XLI, pp. 909-996, 1988.

[5] A. Kumar, D.R. Fuhrmann, M. Frazier, B. Jawerth, "A New Transform For Time-Frequency Analysis," Technical Report WUCS-90-26, Department of Computer Science, Washington University, St. Louis, MO, July, 1990.

[6] R. R. Coifman and M.V. Wickerhauser, "Best-adapted Wave Packet Bases", preprint, Department of Mathematics, Yale University, New Haven, CT, May, 1990.

[7] R. Kasturi, "A Graphics Recognition System - Final Report," Computer Engineering Technical Report TR-90-077, Department of Electrical Engineering, Pennsylvania State University, University Park, Pennsylvania.

[8] Andrew Laine and William Ball, "Applying the Wavelet Decomposition to Character

Recognition", Technical Report WUCS-90-23, Department of Computer Science, Washington University, St. Louis, MO., July, 1990.

[9] R. Kronland-Martinet, "The Wavelet Transform for Analysis, Synthesis, and Processing of Speech and Music Sounds", Computer Music Journal, Vol. 12, No. 4, MIT Press, pp. 11-20, Winter, 1988.

[10] F. Campbell and J. Kulikowski, "Orientation selectivity of the Human Visual System", Journal of Physiology, Vol. 197, pp. 437-441, 1966.

[11] J. Ohya, A. Shio and S. Akamatsu, "A Relaxational Extracting Method for Character Recognition in Scene Images", IEEE Conference on Computer Vision and Pattern Recognition, pp. 424-429, June 5-9, 1988.

[12] I.V. Nagendra and U. G. Gujar, "3-D Objects from 2-D Orthographic Views - a Survey," *Computers & Graphics*, Vol. 12, No. 1, pp. 111-114, 1988.

[13] S. Nishihara and K. Ikeda, "Interpreting Engineering Drawings of Polyhdrons," *Proc. Ninth International Conference on Pattern Recognition*, IEEE Computer Society Press, Nov., 1988.

[14] D. Lysak and R. Kastri, "Interpretation of Line Drawings with Multiple Views," *Proc. Tenth International Conference on Pattern Recognition*, IEEE Computer Society Press, June, 1990.

[15] E.T. Whitaker and M.H. Huhns, "Rule-Based Geometrical Reasoning for the Interpretation of Line Drawings," Applications of Artificial Intelligence III, Proc. SPIE, Vol. 635, pp. 621-627, 1986.

[16] Cohen, "Time-Frequency Distributions-A Review," Proc. of the IEEE, Vol. 77, No. 7, pp. 941-981, July, 1989.

[17] N.G. De Bruijn, "Uncertainty Principles in Fourier Analysis," in *Inequalities*, O. Shisha (ed.), Academic Press, New York, pp. 57-71, 1967.

Appendix II - Auxiliary Figures

Feature Vectors

```
1  blob number in image (1 hex)  *****
43  x centroid
455  y centroid
2684  area
2120  perimiter
-307006.000000  moment M(11)
5057.000000  moment M(20)
224854336.000000  moment M(02)
18836512309248.000000  eccentricity
1674.000000  compactness
0  orientation
90  max grey level
36  min grey level
40  x min coordinate
0  y min coordinate
49  x max coordinate
1008  y max coordinate
2  blob number in image (2 hex)  *****
190  x centroid
1001  y centroid
104  area
86  perimiter
-702.000000  moment M(11)
1987.000000  moment M(20)
1758.000000  moment M(02)
477.240387  eccentricity
71.000000  compactness
-39  orientation
90  max grey level
37  min grey level
180  x min coordinate
994  y min coordinate
196  x max coordinate
1008  y max coordinate
3  blob number in image (3 hex)  *****
208  x centroid
992  y centroid
185  area
148  perimiter
-5888.000000  moment M(11)
3082.000000  moment M(20)
21946.000000  moment M(02)
1923388.875000  eccentricity
118.000000  compactness
16  orientation
90  max grey level
30  min grey level
196  x min coordinate
970  y min coordinate
215  x max coordinate
1008  y max coordinate
4  blob number in image (4 hex)  *****
224  x centroid
1004  y centroid
107  area
84  perimiter
```

Segment 1

Segment 2

Segment 3

Figure A.1. Segment feature vectors of computed geometric properties.

Figure A.2. Sample input for neural network training .