Hexagonal Wavelet Representations for Recognizing Complex Annotations

Andrew F. Laine and Sergio Schuler

Computer and Information Sciences Department
University of Florida
Gainesville, FL, 32611

Abstract

This paper describes a method of pattern recognition targeted for recognizing complex annotations found in paper documents. Our investigation is motivated by the high reliability required for accomplishing autonomous interpretation of maps and engineering drawings. Our approach includes a strategy based on multiscale representations obtained by hexagonal wavelet analysis.

A feasibility study is described in which more than 10,000 patterns were recognized with an error rate of 2.06% by a neural network trained using multiscale representations from a class of 52 distinct patterns. We observed a 21-fold reduction in the amount of information needed to represent each pattern for recognition. These results suggest that high reliability is possible at a reduced cost of representation.

1 Introduction

Fundamental to achieving an autonomous production capability is the development of a *reliable method* for recognizing the characters and symbols contained within a drawing. This paper describes a method of pattern recognition targeted for recognizing complex annotations found in paper documents.

While recent methods of character recognition [1, 12, 14, 16] 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 technology [9, 10]. In particular, the problems of orientation and feature extraction remain unsolved.

We present a method of character recognition that is capable of providing the high reliability needed to make autonomous systems feasible. Our method is embedded into an incremental strategy for recognizing characters based on the multiscale representation of wavelet decompositions [2, 3, 13]. Using wavelets as a set of basis functions, we may decompose an image into a multiresolution hierarchy of localized information at different spatial frequencies. Similar to traditional coarse to fine matching strategies, we attempt first to recognize coarse features within low frequency levels of the wavelet transform. If higher resolution is required to resolve an ambiguity, we may add incrementally to the representation, the finer features of a pattern available at higher frequency levels. Choosing wavelets that are simultaneously localized in both space and frequency, results in a powerful methodology for image analysis. The inner-product of a signal f with a wavelet ψ $(< f, \psi >= \frac{1}{2\pi} < \hat{f}, \hat{\psi} >)$ reflects the character of f within the spatial-frequency region where ψ is localized ($\hat{f}$ and $\hat{\psi}$ are the Fourier transforms of the signal f and the analyzing function ψ). If ψ 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. Both multiorientation and multiresolution are features of the human visual system and wavelet transforms.

Our strategy is to disregard information within the high frequency levels of the basis, and achieve recognition using only spatial-frequency information concentrated within the lower frequency channels. 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 a 4 fold reduction in the number of transform coefficients between descending levels of the hierarchy. Thus, 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 the point of view of information theory, in that the technique converges towards a minimal form of representation without compromising resolution.

We present experimental results, in which over ten thousand *handprinted* characters were recognized with a 2.06% error rate by a neural network [4] trained with dilated wavelet representations (space-frequency coefficients) from a class of 52 distinct alphanumeric and graphical patterns. Our investigation shows that recognition using compressed representations, not only yielded high reliability, but resulted in at least a 21 fold reduction in the amount of information needed to accomplish reliable recognition.

In the next section we describe the context of the recognition problem and motivate our approach. In section 3 we present an overview of our recognition strategy. Finally, section 4 presents a summary and discussion of our results.

2 Motivation and Problem Statement

Manufacturing and utility companies 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. The high reliability and low complexity of our recognition technique is well matched to the requirements of a production quality image understanding system capable of converting paper-based engineering drawings and facilities maps into an electronic form, for storage, retrieval and updates via conventional and geographical databases.

Finding a reliable method of character recognition remains fundamental to accomplishing an autonomous production capability. 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 its lifetime. Invariance to intensity and size is needed since several writing instruments may have been used to draw and maintain the original drawing. Invariance to orientation is needed to handle symbols and text placed at non-horizontal positions. In the next section, we discuss our approach for handling each of these problems.

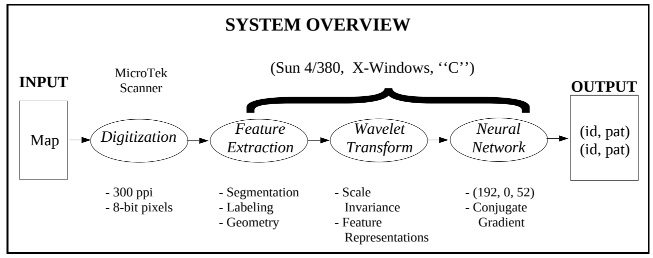


Figure 1: A testbed for pattern recognition.

3 Strategy and Processing Overview

In this section we present a strategy for character recognition exploiting a multiresolution hierarchical basis. We give a general overview of our character recognition system. A formal description of the multiscale basis used in our processing system may be found in [11, 16].

We obtained a number of facility management maps[1] to use as test cases during our investigation. The maps were scanned into digital form, and processed using standard image processing and segmentation routines. In addition, we established a large database of handprinted characters and graphical symbols, comparable to those found in real maps and drawings. This database formed the foundation upon which the accuracy of our character recognition process was evaluated.

In the paragraphs below, we describe our method to extract characters and graphical symbols from paper-based engineering drawings and transform them into an efficient representation for training a neural network to generalize in classification. Our approach, summarized in Figure 1, consists of seven steps:

(1) **Digitize original drawing.** First, each map is digitized at 8-bits per pixel at a resolution of 300 pixels per inch.

(2) **Segment and label connected components.** Next, segmentation is accomplished by labeling 8-connected components having similar grey-level intensity with a unique identifier. We selected a range of grey-level intensities to group similar pixels, for each segmentation. As a preprocessing step, we applied traditional image processing techniques and contrast enhancement (if necessary) before segmentation.

(3) **Compute geometric properties for each labeled segment.** In this step, we computed a set of geometric properties including area, centroid, maxi-

[1] Engineering drawings supplied by Southwestern Bell Technology Resources, Inc., St. Louis, Mo., USA, and McDonnell Douglas Corp., St. Louis, Mo., USA.

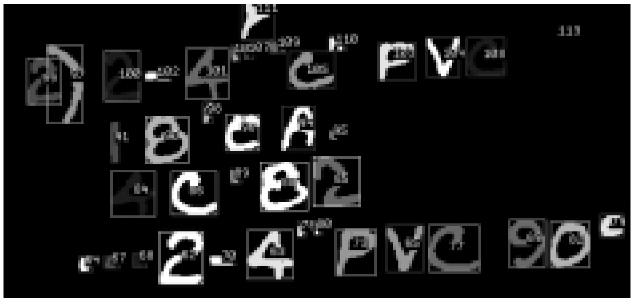


Figure 2: Segmented and labeled components delineated by minimum bounding rectangles.

Figure 3: A shape preserving technique for scale invariance.

mum height and width for all segments. These properties were used later on to separate segments of characters from graphics, establish baselines and to cluster local groups of characters and symbols into meaningful "chunks" of information. In Figure 2, each segment is shown labeled and enclosed within its minimum bounding rectangle.

(4) **Apply geometric constraints to partition segments.** Following techniques developed by Kasturi [5], we used a logical combination of geometric properties computed from the previous step to classify each segment into one of three disjoint partitions: *characters, noise or graphics.* Very small segments were most likely noise (or punctuation), very large segments were most likely pieces of graphics. Segments of a specific height to width ratio and area were coarsely classified as characters, and marked for further processing.

(5) **Scaling and normalization of character segments.** Each character segment is normalized in scale to fit a minimum bounding square of 64 pixels. Here, we introduce a novel method to accomplish *scale invariance.* First we construct a minimum bounding rectangle (MBR) and identify the longest edge, parameter v, as shown in Figure 3. A minimum bounding square (MBS) is then allocated to match the length of

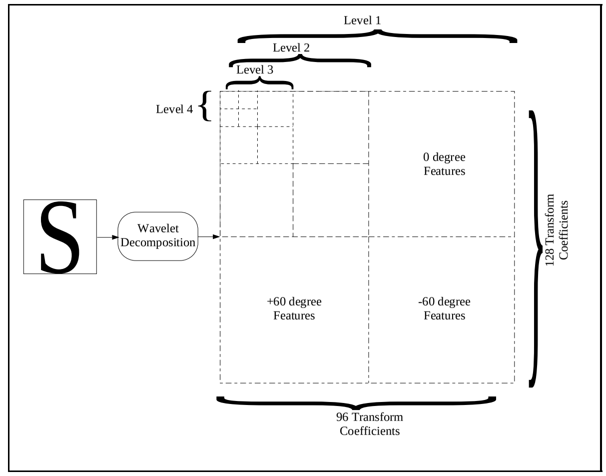


Figure 4: Mapping of features and scale space levels for hexagonal wavelet analysis.

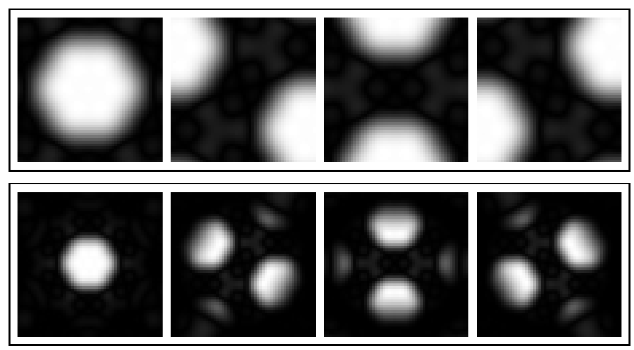
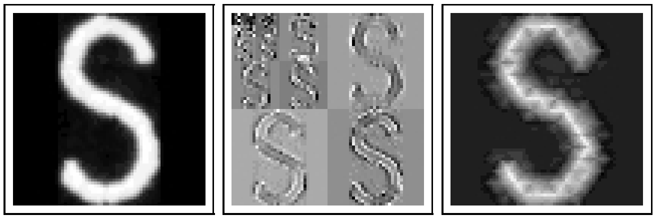


Figure 5: Partitions of the frequency domain resulting from a 2-level multiresolution decomposition using hexagonal filters.

the longest edge of the MBR. Next, the MBR is embedded within the MBS. Once embedded, a character may be shrunk or enlarged without distorting its original shape by the method of bilinear interpolation, as shown in Figure 3.

(6) **Transform characters into a hexagonal multiscale representation.** To map a character segment into a hexagonal sampling lattice we first interpolated its rectangular lattice vertically by a factor of 2 and horizontally by a factor of 3 using bilinear interpolation. The image was then mapped into a hexagonal sampling lattice by the mapping function $M(\mathbf{n}) = (1 + (-1)^{n_1+n_2})/2$, resulting in a 128×96 image matrix. As shown in Figure 4, each segment marked as a character was decomposed by a hierarchical basis into a multiscale representation. In our study, we used a multiresolution decomposition based



(a) (b) (c)

Figure 6: (a) Sample pattern “S”, (b) Magnitude of the wavelet coefficients for a four level decomposition, (c) Reconstruction from level 4 coefficients alone.

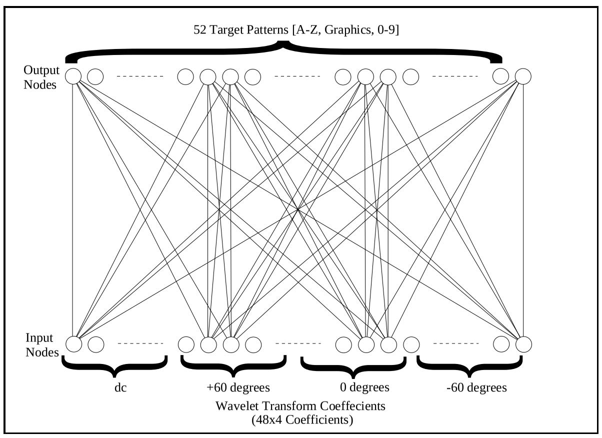


Figure 7: A network topology for multiscale character recognition.

on hexagonal wavelets [11, 16]. The hexagonal wavelet transform decomposes the information within each character segment into a multiscale space-frequency representation with orientation selectivity (0, +60 and −60 degree bands). Figure 5 shows the partition of the frequency domain resulting from a 2-level multiresolution decomposition of hexagonal filters. In our investigation, we were able to reduce the number of transform coefficients needed to discriminate a class of 52 distinct patterns by a factor of 21 (4096 original vs. 192 compressed). Figure 6 shows a four level hexagonal wavelet decomposition for the pattern “S” and its reconstruction from level 4 coefficients alone. Notice that these coefficients ($8 \times 6 \times 4$) are sufficient to preserve the fundamental shape of the character. Figure 7 shows how level 4 coefficients of a hexagonal multiscale representations were mapped onto the topology of a single layered neural network.

(7) **A simple neural network topology.** Next, the low frequency components of the wavelet decomposition are applied to train a neural network to recognize the 52 alphanumeric characters [A-Z, Graphics, 0-9].

Database of Sample Patterns				
Patterns	Number of Distinct Patterns	Number of Samples		
		Training	Testing	Total
0-9	10	1,360	680	2,040
Graphics	16	2,176	1,088	3,264
A-Z	26	3,536	1,768	5,304
All Above	52	7,072	3,536	10,608

Table 1: Distribution for database of sample patterns.

We used the conjugate gradient method for training a single-layer neural network [4]. Since we had reduced the representation needed for each character and symbol, our neural network required only 192 input units (one for each transform-coefficient), rather than 4096 input units, that would had been needed to train a network using the original representation. The rest of the network consisted of zero hidden units and 52 [A-Z, Graphics, 0-9], 26 [A-Z] or 10 [0-9] output units.

Scaling, translation, and multiple fonts are handled quite well by the procedures described above. However, the steps described in the previous section assume that characters are mostly horizontal in orientation.

Our approach for handling non-horizontal characters is unique. First we identify clusters of characters 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 base line. Thus all characters are processed uniformly. This approach has the advantage in that it is far less complicated and more reliable than classifiers that attempt to recognize each character in its many possible orientations.

4 Summary of Results and Discussion

We have presented a method of character recognition based on spatial-frequency representations of multiresolution wavelet transforms. We have built a database of alphanumeric and graphical patterns of handwritten samples collected from engineers and scientists. The database consisted of 52 distinct patterns and contained over 10,000 samples. Table 1 summarizes the distribution of patterns in our database. These samples were comparable in quality to those found in real engineering documents (e.g. compare Figures 2 and 8).

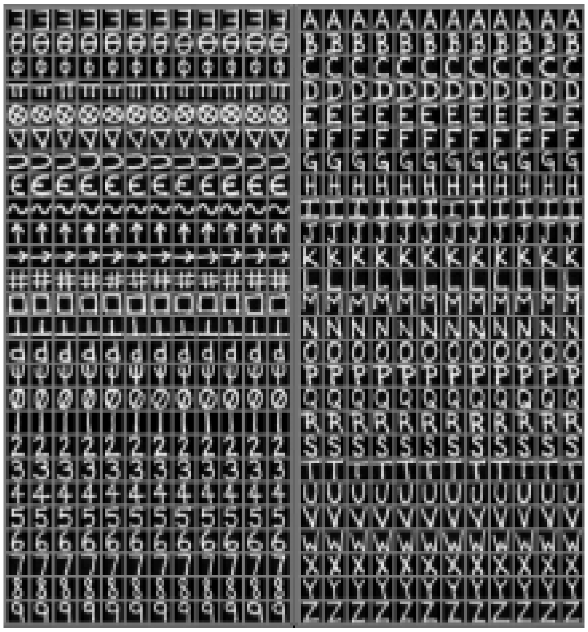


Figure 8: Segmented and scaled characters.

In our evaluation of wavelet representations for recognition, we trained and tested subsets of our database using distinct neural network topologies. Table 2 outlines the classification cases considered in our investigation in increasing order of complexity. Our experiments show that representations obtained at level 4 (shown in Figure 6) were sufficient for accomplishing highly reliable recognition. This results from the superior performance observed on a neural network classifier when trained exactly with scaled, +60, 0, and −60 degree features. We have included a comparison between the performance in classifying patterns using hexagonal wavelet representations at level 4 and previous work [10] using traditional rectangular wavelet representations at level 3. Notice that hexagonal wavelets representations yield more reliable results than their rectangular wavelet counterparts for the same network topologies. We attribute this performance advantage to the fact that our hexagonal wavelet representation partitions orientations into three bands of 60 degrees covering equally the frequency domain, *favoring no particular orientation* and resolving the *mixed orientation problem* inherent to traditional (tensor product) rectangular wavelet representations.

Quantitatively, we reduced the number of coefficients required to accomplish recognition by more than 21 fold; from 64×64 (original input pattern) to $8\times6\times4$ (transform coefficients for level 4). Thus only 192 input units were needed to configure the network, rather than 4096 which would have been needed to train a network using the original representation. Next, we compare our results against several existing character recognition techniques [6, 7, 8, 15].

Performance Evaluation				
Cases	Hexagonal Wavelets		Rectangular Wavelets	
	Number of Errors	Error Rate	Number of Errors	Error Rate
0-9	6	0.88%	11	1.61%
A-Z	20	1.13%	26	1.47%
A-Z, Graphics, 0-9	73	2.06%	93	2.60%

Table 2: Performance evaluation for three classification cases: [0-9], [A-Z] and [A-Z, Graphics, 0-9] for hexagonal wavelet representations at level 4 and rectangular wavelet representations at level 3 (scaled, horizontal, and vertical features only).

Complexity Evaluation			
Cases	Network Configuration	Number of Epochs for Hexagonal Wavelets	Number of Epochs for Rectangular Wavelets
0-9	192-0-10	207	255
A-Z	192-0-26	376	224
A-Z, Graphics, 0-9	192-0-52	494	453

Table 3: Complexity evaluation for classification cases shown in Table 2.

Khotanzad and Lu [8] trained and tested a neural net classifier using invariant Zernike features. Their database consisted of a total of 624 samples, for 26 distinct patterns [A-Z], or 24 samples per pattern. Each set of 24 samples was constructed by varying the scale, translation and orientation of 4 distinct instances of each pattern. Note that in this study each pattern was normalized with respect to scale and translation by using regular moments [6], while orientation invariance was accomplished through Zernike features [7]. Thus the resulting database actually consisted of only 104 distinct samples. In comparison, our database contained 204 distinct instances per pattern resulting in 5304 distinct samples for the patterns [A-Z]. The variations between patterns in both databases are similar.

The performance obtained by Khotanzad and Lu [8] training a neural net classifier configured as a multi-layer perceptron (47-50-26) resulted in 100% classification accuracy for training and testing sets each consisting of 312 samples (52 distinct classes). This result required 500 epochs over the training set. However, our method also reported 100% accuracy for training and testing sets on a database of the same size (624 samples).

As shown in Tables 2 and 3, for the classification case [A-Z], our neural net converged after 376 epochs of training 3,536 patterns, and yielded an error rate of 1.13% on testing 1,768 patterns. Configured in a *single-layer* topology (192-0-26), our network converged within 9 hours executing on a Sun 10/30 with 64 megabytes of memory. These results make high reliability possible at a *reduced cost* of representation and training.

Recently, Eduard Säckinger *et. al* [15] trained a neural network of 136,000 connections for recognition of handwritten digits [0-9] using a mixed analog/digital neural network chip (VLSI). The neural network classifier was a five-layer network consisting of 400 inputs (corresponding to a 20×20 pixel image), and 10 outputs. No preprocessing steps were employed but the connections in the first four layers of the network were spatially constrained. In their study a neural net (simulated on a Sun workstation) achieved an error rate of 4.9% on a test set obtained from segmented, handwritten ZIP codes extracted from real U.S. mail correspondence.

By comparison, we observed that a single layer network accomplished recognition of handprinted numbers using orthogonal wavelet representations with a 0.88% error rate. In addition, the reduced complexity provided by a simplified interconnection topology is attractive for low-cost (high reliability) VLSI implementation.

Our results demonstrate that wavelet representations are an efficient basis for neural net classification since feature vectors cluster into nearly linearly separable classes. This has been shown experimentally by the reliable performance obtained using a neural net classifier without a hidden unit layer.

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.

References

[1] D.J. Burr, "Experiments on Neural Net Recognition of Spoken and Written Text," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 36(7), pp. 1162–1168, 1988.

[2] R.R. Coifman, M.V. Wickerhauser, "Entropy-Based Algorithms for Best Basis Selection," IEEE Transactions on Information Theory, Vol. 38(2), pp. 713–718, 1990.

[3] I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets," Communications on Pure and Applied Mathematics, Vol 41, pp. 909–996, 1988.

[4] B. Kalman, "Super Linear Learning in Back Propagation Neural Nets," Department of Computer Science, Technical Report WUCS–90–21, Washington University, St. Louis, Missouri, 1990.

[5] R. Kasturi, S.T. Bow, W. El-Masri, J. Shah, J.R. Cattiler, U.B. Mokate, "A System for Interpretation of Line Drawings," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12(10), pp. 978–991, 1990.

[6] A. Khotanzad, J.H. Lu, "Object Recognition Using a Neural Network and Invariant Zernike Features," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, California, pp. 200–205, 1989.

[7] A. Khotanzad, Y.H. Hong, "Invariant Image Recognition by Zernike Moments," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. (12)5, pp. 489–497, 1990.

[8] A. Khotanzad, J.H. Lu, "Classification of Invariant Image Representations Using a Neural Network," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. (38)6, pp. 1028–1038, 1990.

[9] A. Laine, W. Ball, A. Kumar, "A Multiscale Approach for Recognizing Complex Annotations in Engineering Documents," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Lahaina, Maui, Hawaii, pp. 749–750, 1991.

[10] A. Laine, S. Schuler, V. Girirsh, "Orthonormal Wavelet Representations for Recognizing Complex Annotations," Machine Vision and Applications, Vol. 6, pp. 110–123, 1993.

[11] A. Laine, S. Schuler, "Hexagonal wavelet processing of digital mammography," Medical Imaging 1993, Part of SPIE's Thematic Applied Science and Engineering Series, Newport Beach, California, February 14–19, 1993.

[12] H.F. Li, R. Jayakumar, M. Youssef, "Parallel Algorithms for Recognizing Handwritten Characters Using Shape Features," Pattern Recognition, Vol. 22(6), pp. 641–652, 1989.

[13] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," IEEE Transactions on Pattern Analysis and Machine Intelligence 11(7): 674–693, 1989.

[14] J. Ohya, A. Shio, S. Akamatsu, "A Relaxational Extracting Method for Character Recognition in Scene Images," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Ann Arbor, Michigan, pp. 424–429, 1988.

[15] E. Säckinger, B.E. Boser, J. Bromley, Y. LeCun, L.D. Jackel, "Application of the ANNA Neural Network Chip to High-Speed Character Recognition," IEEE Transactions on Neural Networks Vol. 3(3), pp. 498–505, 1992.

[16] E.P. Simoncelli, E.H. Adelson, "Non-separable Extensions of Quadrature Mirror Filters to Multiple Dimensions," Proceedings of the IEEE, Vol. 78(4), pp. 652–663, 1990.