A MULTISCALE APPROACH FOR RECOGNIZING COMPLEX ANNOTATIONS IN ENGINEERING DOCUMENTS

Andrew Laine
Computer and Information Sciences Department
University of Florida
Gainesville, Florida 32611-2024

William Ball and Arun Kumar
Department of Computer Science
Washington University
Saint Louis, Missouri 63130-4899

ABSTRACT

This paper describes a novel method of character recognition targeted for complex annotations found in engineering documents. A feasibility study is described in which characters extracted from engineering drawings were recognized without error from a class of 36 distinct alphanumeric patterns by a neural network classifier trained with multiscale representations. We present an incremental strategy for resolution which relies upon the continuity between hierarchical levels of a novel multiscale decomposition. We observed a 16-fold reduction in the amount of information needed to represent each character for recognition. These results suggest high reliability 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. Problems of orientation (recognizing text placed non-horizontally) and feature extraction (separation of text from graphics) complicate the recognition task.

Some important contributions to engineering drawing interpretation are summarized in a survey paper by Nagendra and Gutar [1]. In addition recent works by Hishihara and Ikeda [2], and Kasturi [3] describe significant advances towards autonomous interpretation of maps and engineering drawings. However, highly reliable methods for recognizing characters and symbols remain a fundamental problem for achieving an autonomous production capability.

We have developed a novel incremental strategy for pattern recognition that utilizes continuity (bijection) between hierarchical levels of a multiresolution space-frequency decomposition called the Frazier-Jawerth transform (FJT). Frazier-Jawerth transforms [4,5] are closely related to wavelet decompositions [6] and are more attractive than traditional hierarchical multiresolution decompositions because they are linear, continuous, and continuously invertible.

Several hundred characters extracted from real engineering drawings were recognized without error by a neural network trained with dilated representations (FJT coefficients) from a class of 36 distinct alphanumeric patterns. Our investigation suggests that high reliability may be obtained using such efficient (nonredundant) representations.

In the sections below, we describe our incremental strategy for multiscale recognition and present a system overview, including a brief description of the decomposition and analyzing functions used in our investigation.

2. An Incremental Strategy for Multiscale Recognition

Similar to traditional coarse to fine matching strategies, we first attempt to recognize coarse features within low space-frequency levels of the 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 space-frequency levels. There is a 4 fold reduction in the number of transform coefficients within each level of the multiresolution hierarchy. Thus, accomplishing recognition using information available at the lower levels of the hierarchy, requires fewer transform coefficients to represent each pattern. This is desirable from an information theoretic point of view in that the technique converges towards a minimal (efficient) form of representation without compromising discrimination.

Our approach is motivated in part by multiorientation and multiresolution mechanisms of the human visual system. Choosing analyzing functions that are well localized in space, results in a powerful methodology for image analysis. The inner-product of a signal x with an analyzing function 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. In contrast to ad-hoc approaches, the incremental strategy presented in this paper promises a practical solution embedded in a unified mathematical theory.

3. System Overview

Invariance to font style, intensity, scale, and orientation is accomplished and presented in [7]. Below, we describe the preprocessing steps used to extract annotations, and decompose them into compact representations used to train a neural network classifier.

Each drawing is first digitize and segmented into labelled components, where each component is bounded by its minimum rectangle [3] as shown in Figure 1. In addition, several geometric properties (constraints) are computed for each segment and used to classify segments into three disjoint partitions: characters, noise or graphics. Each character segment is normalized in scale to fit a minimum bounding square of 64 pixels, by first identifying the longest edge of its minimum bounding rectangle (MBR). 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 its MBS so that each character may be shrunk or enlarged without distorting its original shape.

Next, we decompose the information within the bounding square into a multiscale representation. The FJ decomposition of a signal f is defined in [4], $f : \mathbf{R}^n \rightarrow \mathbf{C}$ in $L^2(\mathbf{R})$,

CH2983-5/91/0000/0749/$01.00 © 1991 IEEE

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

The signal f is written as a weighted sum of sum of *synthesizing functions* from the set $S = \{\Psi_{mk},\psi_{vk}\}_{v,k}$, where m is a fixed integer, $v > m$, and $k \in Z^n$. The functions Φ_{mk} in S are the translates (in R^n) of the single "parent" function Φ_{m0}, while ψ_{vk} are translated-and-dilated versions of a single function ψ. The FJT of a signal f is the countably-infinite sequence $Tf = (\langle f,\Phi_{mk}\rangle, \langle f,\phi_{vk}\rangle)_{v,k}$ of inner-products of f with *analyzing functions* from the set $A = \{\Phi_{mk}, \phi_{vk}\}$.

In our study, we used cosine-of-log functions for both analyzing and synthesizing functions, because they were relatively simple to design. 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 2, so that:

$$\hat{\Theta}_m + \sum_{v=m+1}^{\infty}\hat{\theta}_v \equiv 1, \ \forall\omega\in R. \quad (2)$$

As shown in Figure 3 (rightmost two columns) our analyzing function (ϕ) has bounded support in the frequency domain, and *localized* support in the time domain.

In (2) we assume that the *support* $\hat{\Theta} \subseteq \{||\omega|| \le \pi\}$, and *support* $\hat{\theta} \subseteq \{\frac{\pi}{4} \le ||\omega|| \le \pi\}$. Thus

$$f = f\hat{\Theta}_m + \sum_{v=m+1}^{\infty} f\hat{\theta}_v, \ \forall\omega\in R. \quad (3)$$

We used a simple construction where the "window-function" $\hat{\theta}$ was chosen to be a raised cosine pulse:

$$\hat{\theta}(\omega) \equiv \begin{cases} \frac{1}{2}(1-\cos(\pi\log_2||\omega||)) & , \frac{\pi}{4} \le ||\omega|| \le \pi \\ 0 & , \text{otherwise} \end{cases} \quad (4)$$

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 (2) over the frequency interval $[-\pi, \pi]$. The functions $\hat{\phi}$ and ψ were chosen to be the square-root of $\hat{\theta}$, and $\hat{\Theta}$ and $\hat{\theta}$ were factored such that the DC cap functions $\hat{\Phi} = \hat{\Psi}$ and $\hat{\phi} = \hat{\psi}$ (making the analyzing and synthesizing functions the same).

4. Summary of Results

We have presented a novel method of character recognition based on a multiscale space-frequency transformation, closely related to the 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 spatial and frequency domains.

Experimentally, we observed a 16-fold reduction in the amount of information needed to represent a class of 36 alphanumeric patterns [A-Z,0-9]. For most characters sampled, representation at dilation level Φ_{-3} 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). As a result, our classifier (a two-layer 64-20-36 neural network, trained by backpropagation) required only 64 input units in its configuration rather than 4096. In addition, less time was required for training. These results suggest high reliability at a *reduced cost* of representation.

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

5. References

[1] 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.

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

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

[4] M. Frazier and B. Jawerth, "The ϕ-Transform and Applications to Distribution Spaces," in *Function Spaces and Applications*, M. Cwikel et al. (eds.), Springer Lecture Notes in Mathematics, no. 1302, pp. 223-246, 1988.

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

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

[7] Andrew Laine, William Ball and Arun Kumar, "A Multiscale Approach for Recognizing Complex Annotations in Engineering Documents", Technical Report WUCS-90-23, Washington University, St. Louis, MO., July, 1990.

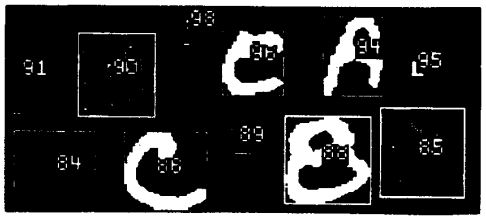

Figure 1.

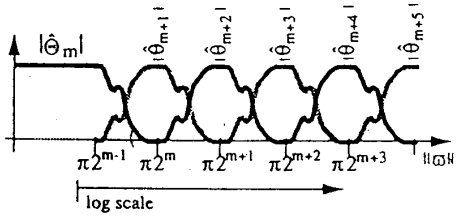

Figure 2.

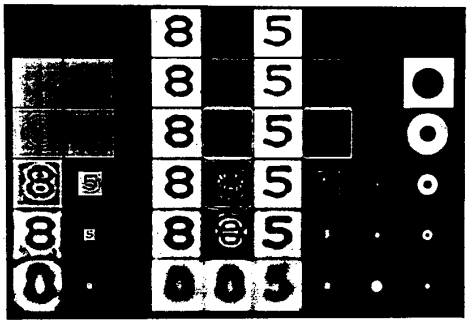

Figure 3.