of generalization performance (Denker *et al.* 1987; Baum and Haussler 1989) The present system performs slightly better than the previous system This is remarkable considering that much less specific information about the problem was built into the network. Furthermore, the new approach seems to have more potential for improvement by designing more specialized architectures with more connections and fewer free parameters. [1]

Waibel (1989) describes a large network (but still small compared to ours) with about 18,000 connections and 1800 free parameters, trained on a speech recognition task. Because training time was prohibitive (18 days on an Alliant mini-supercomputer), he suggested building the network from smaller, separately trained networks. We did not need such a modular construction procedure since our training times were "only" 3 days on a Sun workstation, and in any case it is not clear how to partition our problem into separately trainable subproblems.

5.2 DSP Implementation. During the recognition process, almost all the computation time is spent performing multiply accumulate operations, a task that digital signal processors (DSP) are specifically designed for We used an off-the-shelf board that contains 256 kbytes of local memory and an AT&T DSP-32C general purpose DSP with a peak performance of 12 5 million multiply add operations per second on 32 bit floating point numbers (25 MFLOPS). The DSP operates as a coprocessor; the host is a personal computer (PC), which also contains a video acquisition board connected to a camera.

The personal computer digitizes an image and binarizes it using an adaptive thresholding technique. The thresholded image is then scanned and each connected component (or segment) is isolated. Components that are too small or too large are discarded; remaining components are sent to the DSP for normalization and recognition. The PC gives a variable sized pixel map representation of a single digit to the DSP, which performs the normalization and the classification.

The overall throughput of the digit recognizer including image acquisition is 10 to 12 classifications per second and is limited mainly by the normalization step. On normalized digits, the DSP performs more than 30 classifications per second.

6 Conclusion

We have successfully applied backpropagation learning to a large, real-world task. Our results appear to be at the state of the art in digit recognition. Our network was trained on a low-level representation of

[1] A network similar to the one described here with 100,000 connections and 2600 free parameters recently achieved 9% rejection for 1% error rate That is about 30% better than the best of the hand-coded-kernel networks