# Pen-based Visitor Registration System (PENGUIN)

N. Matić,* D. Henderson, Y. Le Cun, Y. Bengio
AT&T Bell Laboratories
Holmdel, NJ 07733

**Abstract**

*We describe a new electronic Pen-based Visitors Registration System (PENGUIN) whose goal is to expand and modernize the visitor sign-in procedure at Bell Laboratories. The system uses a pen-interface (i.e. tablet-display) in what is essentially a form filling application. Our pen-interface is coupled with a powerful and accurate on-line handwriting recognition module [1, 2]. A database of AT&T employees (the visitors' hosts) and country names is used to check the recognition module outputs, in order to find the best match.*

*The system provides assistance to the guard at one of the guard stations in routing visitors to their hosts. All the entered data are stored electronically. Initial testing shows that PENGUIN system performs reliably and with high accuracy. It retrieves the correct host name with 97% accuracy and the correct visitors citizenship with 99% accuracy. The system is robust and easy to use for both visitors and guards.*

## 1 Introduction

The ways in which people communicate with computers are changing very fast. More and more different forms of access (e.g. speech, keyboards, mouse, pen-based interfaces) are available for different type of applications. The pen-interface is a type of "electronic paper" that consists of a transparent drawing surface overlaid on top of the tablet display. It accurately captures x-y coordinates of the pen-tip movement. ("electronic ink" displays the trace on the screen surface). Pen-based interfaces have a potential for becoming the vehicle for an increasing range of applications since more and more tasks are becoming computer-based. They are appealing since they support a "natural" form of human - computer interaction. They are likely to become more popular since we expect increasing number of untrained users to interact with computers. However, their use is justified only when handwritten input is a convenient way of entering data, and when they are combined with handwriting recognition technology that enables immediate conversion of the input information into the coded information convenient for computer storing and processing.

Handwriting recognition is a difficult problem. Though much research has been done both in industry and in research labs, highly accurate recognition for general tasks is hard to achieve, mainly because of high degree of variability inherent to the task [3].

*The author is now with Synaptics Inc. San Jose

Keeping this in mind, to succeed, an application that uses a pen-based interface and handwriting recognition must tightly restrict the task domain in order to keep complexity low and recognition accuracy high. Typically it must also make use of some language model or dictionary. Also, the processing speed has to be adequate for real-time use.

In this paper we describe an electronic Pen-based Visitor Registration System (PENGUIN) that will modernize and expand the capabilities of the current visitors sign-in procedure at Bell Laboratories. For a visitor, the system provides a non-intrusive procedure that is essentially equivalent to filling out the current paper form, but having the paper replaced by an electronic tablet. The PENGUIN system uses handwriting recognition technology and database search in order to provide assistance to the user (in this case a guard at one of the guards stations) in routing Bell Lab's visitor to her/his destination host office. All the entered data is stored on disk.

By using the visitor's handwritten input, the system is making an attempt to recognize some of the entered data, like visitor's citizenship, and host's name.

The system uses a recognition engine, trained on isolated uppercase characters and digits. The recognizer uses a coarse-coded spatial representation called AMAP [1] in combination with multilayer *convolutional* neural networks [4, 5] previously successfully used in optical character recognition. The raw performance of the recognizer is very high, which makes it possible to obtain extremely high accuracy after the personnel and countries dictionary search. Initial testing indicates that the system performs with very high accuracy and leads to unique host candidates in most cases.

## 2 Current Visitor Sign-in Procedure and Proposed Solution

Current visitor sign-in procedure consists of the following steps: the following:

- The visitor fills-out paper form, with his name, his company information, citizenship, and host's name and optionally host's phone number.
- The visitor has to write his/her name on the badge form, and the guard stamps it with the current date.
- The guard has to enter the host's name specified on the form into the computer in order to extract



1058-6393/95 $4.00 © 1995 IEEE

information about the host. In the case when host's name is misspelled or imprecisely written, this step has to be repeated.

- Finally the guard has to call the host.
- No information is stored on the disk.

In order to modify and modernize sign-in procedure, we have identified some of the requirements that the system has to provide.

It has to be non-intrusive towards the visitor. A visitor has to fill a form that is essentially equivalent to paper form, on the electronic tablet device, in her/his own handwriting. No knowledge of the computing is assumed.

For the guard the system has to act as a "personal assistant" . Based on the visitor's input data entered in "natural", handwritten format the system has to convert data into a format used by the computer. It also has to offer useful information to the guard (e.g. provide the host's full name, phone number and room number) and optionally perform other actions such as host phone dial-up etc.

The system has to be able to receive the guards feedback, when it gives wrong directions (or no directions at all).

The system has to handle misspelling and imprecisely entered data robustly. All entered data must be stored on the disk. Furthermore, we would like to store as much as possible of the input data entered in handwritten form in ASCII format.

## 3 The PENGUIN System

The PENGUIN System was designed based on above requirements. It consist of two pen-based interfaces (e.g. WACOM tablets), two terminals (one for each guard), one workstation (Sun Sparc-10) and one printer. The architecture of the system is given on Figure 1.

Each guard terminal has three windows (see figure 2). In the left window a bit-map of the tablet form is displayed. Two windows on the right are: a dialog window; where the results of the recognition process are displayed, and a window for the existing Employee Verification System (EVS), currently in use on the guard desk. When a employee forgets his/her pass, the EVS system is queried, and a daily pass is issued.

The complete visitor sign-in procedure on the PENGUIN system consist of several steps:

- The visitor has to fill-in the Visitor Registration Form that is displayed on the tablet, using electronic pen.
- In the left window of the corresponding guard terminal (see Figure 2), the content of the visitor register form entered on the tablet is displayed.
- The visitor badge is automatically printed, with visitor's name printed in his/hers own handwriting (as entered on the tablet) and the date.

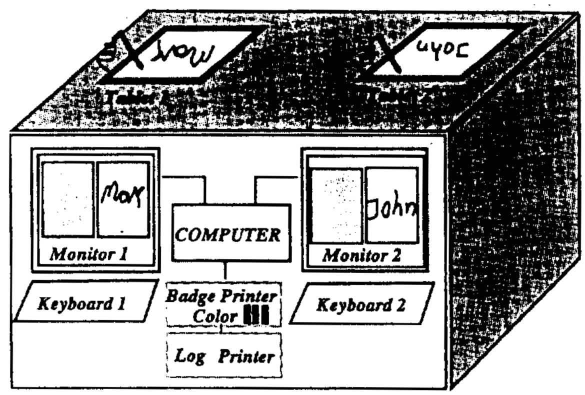


Figure 1: Visitor Registration System at guard station

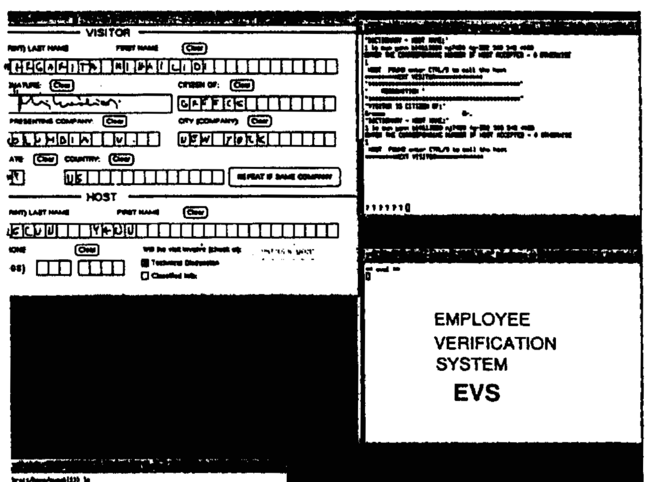


Figure 2: Guards terminal display. The two windows on the right are: Dialog window, where the results of the recognition process are displayed, and Employee Verification System (EVS). On the left, a copy of the tablet form is displayed.

```
"DICTIONARY - HOST NAME:"
1 le cun yann b10113590 nj7460 4g-332 908 949 4038
ENTER THE CORRESPONDING NUMBER IF HOST ACCEPTED - 0 OTHERWISE
1
 HOST  FOUND enter CTRL/D to call the host
*********NEXT VISITOR**************
"##############################################"
"    RECOGNITION "
"##############################################"
"VISITOR IS CITIZEN OF:"
Greece                          Gr.
"DICTIONARY - HOST NAME:"
1 le cun yann b10113590 nj7460 4g-332 908 949 4038
ENTER THE CORRESPONDING NUMBER IF HOST ACCEPTED - 0 OTHERWISE
1
 HOST  FOUND enter CTRL/D to call the host
*********NEXT VISITOR**************
```

Figure 3: Dialog window. See also Figure 2. The results of the recognition are displayed here. The guard then has to accept or reject the host.

- In the dialog window of the guard's terminal (see Figures 2 and 3) the results of handwriting recognition are displayed in real-time. First a citizenship is displayed. Most often, the citizenship suggestion is unique, but sometimes, more than one country is displayed. Secondly, the host name and phone are displayed. Again, most often unique name is given, but in general up to ten tentative hosts are displayed. The guard has to accept or reject the host name. If the host is accepted, the PENGUIN system automatically dials host's phone number.
- All data entered by a visitor, are stored on the disk (some in ASCII format). A browser for later inspection of this data has been developed.

# 4 Recognition

On-line handwriting recognition is a key component in a large number of nontrivial applications that make use of pen-based interfaces. The recognition of handwritten characters from a pen trajectory on a digitizing surface is often carried out in the time domain. Trajectories are normalized, and local geometrical or dynamical features are extracted. The recognition techniques vary widely depending upon the features chosen, the way they were extracted etc. [6, 7]. Unfortunately, despite many advantages, approaches that makes use of the dynamic of handwriting tend to be sensitive to stroke ordering. In addition, global geometric characteristics, such as whether a stroke crosses another stroke drawn at a different time, are difficult to extract.

For the handwriting recognition, we use a novel type of recognizer [1, 2]. The recognizer makes use of a special representation scheme, called AMAP, that preserves the pictorial nature of the handwriting, in combination with a Multi-Layer Convolutional Neural Networks (MLCNN) [4, 5]. The recognizer was trained to be writer independent.

Current implementation requires that a tablet form is filled-in using upper-case characters and digits. We make use of the corresponding neural network trained classifier on the AMAP input representation of each digit and upper-case character. In our case, segmentation is trivial since characters are entered in boxes.

The choice of this particular recognizer was based on several advantages:

- The recognizer is not sensitive to stroke ordering or writing speed.
- It is suitable for any type of handwriting (uppercase, lowercase, symbols, and cursive).
- The raw accuracy is high.
- AMAP's can be computed for complete words without segmentation [2], which is important for the future development of the PENGUIN system, where the integrated segmentation and recognition will be necessary.

## 4.1 AMAP

Since the intent of the writer is to produce a legible *image* of the character, it is natural to preserve as much of the pictorial nature of the signal as possible, while at the same time exploiting the sequential information in the trajectory. In the AMAP representation scheme, pen trajectories are represented by low-resolution images in which each picture element contains information about the local properties of the trajectory. More generally, an AMAP can be viewed as a function in a multidimensional space where each dimension is associated with a local property of the trajectory, say the direction of motion $\theta$, the $X$ position, and the $Y$ position of the pen. The value of the function at a particular location $(\theta, X, Y)$ in the space represents a smooth version of the "density" of features in the trajectory that have values $(\theta, X, Y)$ (This is in the spirit of the generalized Hough transform). An AMAP is a multidimensional array (say 4x10x10) obtained by discretizing the feature density space into "boxes". Each array element is assigned a value equal to the integral of the feature density function over the corresponding box. For more details on how this is computed in practice see reference [1].

## 4.2 Convolutional Neural Networks

Image-like representations such as AMAPs are particularly well suited for use in combination with Multi-Layer Convolutional Neural Networks (MLCNN) [4, 5]. MLCNNs are feed-forward neural networks whose architectures are tailored for minimizing the sensitivity to translations, rotations, or distortions of the input image. They are trained with a variation of the Back-Propagation algorithm.

The units in MCLNNs are only connected to a local neighborhood in the previous layer. Each unit can be seen as a local feature detector whose function is determined by the learning procedure. Insensitivity to local transformations is built into the network architecture by constraining sets of units located at different places to use identical weight vectors, thereby forcing them to detect the same feature on different parts of the input. The outputs of the units at identical locations in different feature maps can be collectively thought of as a local feature vector. Features of increasing complexity and globality are extracted by the neurons in the successive layers. This weight-sharing technique has interesting side effect. The number of free parameters in the system is greatly reduced since a large number of units share the same weights. In our application we use MLCNNs where a single character is shown at the input, and have a single set of outputs.

The networks we use have 5 layers arranged as follows: layer 1: convolution with 8 kernels of size 3x3, layer 2: 2x2 subsampling, layer 3: convolution with 25 kernels of size 5x5, layer 4 convolution with 84 kernels of size 4x4, layer 5: 2x2 subsampling. The subsampling layers are essential to the network's robustness to distortions. The output layer is an 84-dimensional vector. The target output configuration for each character class was chosen to be a *bitmap* of the corresponding character in a standard 7x12 (=84) pixel font. Such a code facilitates the correction of confusable characters by the postprocessor.

Experiments performed on isolated uppercase characters and digits reported in [2], shows the following performance on the independent test sets for the writer independent case. On a test set of size 9,122 uppercase characters the error rate was 4.9%. On a test set of size 2,938 digits, the error rate was 1.25%. Two separate MLCNN were trained for uppercase characters and digits.

### 4.3 Dictionary Search

The recognizer generates character sequence hypothesis for each input word. Each character in a sequence is the one with the highest recognition score. This tentative answer has to be checked against dictionary of allowed words.

There are two dictionaries in our system. The first one is a small size dictionary of countries. The second dictionary is a database of AT&T employees at particular AT&T location. The size of that dictionary can be several thousand entries. Each entry consists of employee last name, first name or first initial (optionally middle name or middle initial), room location, phone, e-mail etc.

The recognizer provides a word (or words) for a country name and typically two words for host's first and last and name. If a phone number has been entered, an additional word consisting of a digit sequence is generated.

There are different sources of errors in tentative answers provided by the recognizer. A visitor may not enter the exact spelling of the country or host name, or may not provide the exact first name of the host or provides only the last name and first initial. The recognition process itself also introduces errors.

For the above reasons a dictionary check is essential for good accuracy.

In our application, we found that an approximate string-matching technique [8], that allows for robust and fast searching of the text in the presence of errors, is sufficient to obtain high accuracy. This approximate string matching algorithm is an extension of the exact string matching algorithm of Baeza-Yates and Gonnet [9]. We use it both for the country database and corporate personnel database search. The tools for approximate search on UNIX systems are in the form of *agrep* command.

## 5 Testing Results

We have tested the Penguin system in in AT&T Bell Laboratories, Holmdel. Conditions were quite similar to those expected at the guard station. We had 105 participants. Each participant had to fill the form displayed on the tablet. They were instructed to fill mandatory entries in the form, such as visitors name and citizenship and host name. In most cases they did not fill the information on host's phone number.

The PENGUIN system makes an attempt to recognize visitor's citizenship and host name. We have measured the performance of the recognizer both on the character level and word level. We have noticed that there is a significant difference in the performance results on the *character level* in PENGUIN test experiments ( "walk in" accuracy), and the performance results reported in [2] on the independent test set (see section 5.2). The walk-in accuracy on the character level is much worse than on the test set (section 5.2) . One explanation would be that the pool of writers in our cafeteria test reflect the worst-case scenario (i.e. people who see and use the tablet for the first time) while a typical test data (for testing the recognition engine) consists of handwriting samples collected from a pool of more experienced tablet users.

Out of 105 participants the legal number of entries were 100 for visitor's citizenship (example of non-legal entry would be "The World" or non-existing country). We have measured the performance of the recognizer on the character level. On the character level the error rate was 8%. The word-phrase error was 21%. (word-phrase in this case consists of country name that is either single word or phrase, i.e. more than one word). After the dictionary search, the word-phrase error goes down to 1.01%.

Out of 105 participants 103 entered a valid host name. (Example of a illegal host is the name of a person that is not employee of Bell Laboratories). On the character level error rate was 9%. On the word-phrase level (word-phrase is either last name, last name and first name, or last name and first initial) the error rate was 53%. After dictionary search error went down substantially to 3.06%.

## 6 Conclusions

We have described the PENGUIN system, a pen-based Visitors Registration System that provides efficient and fast visitor sign-in procedure at the guard stations at Bell Laboratories. The PENGUIN system employs a convolutional neural network recognizer followed by database search. Excellent "walk-in" accuracy on the word (phrase) level has been obtained in our testing where most of the participants where the first time users of any type of pen-based user interfaces.

## 7 Acknowledgments

The authors would like to thank Dave Alles, for generous support for this project. Special thanks to Dave Graepel and Tim Clayton for their help and constructive suggestions. Finally, thanks to all the neural network research group headed by L. Jackel at Bell Labs, Holmdel, for their suggestions regarding this work. All the simulations were carried out using SN2.6, a neural network simulator package developed by Neuristique.

## References

[1] Y. Le Cun, Y. Bengio, D. Henderson, A. Weisbuch, H. Weissman, and Jackel L. On-line handwriting recognition with neural networks: spatial representation versus temporal representation. In *Proc. International Conference on handwriting and drawing.* Ecole Nationale Superieure des Telecommunications, Paris, 1993.

[2] Y. Bengio, Y. Le Cun, and D. Henderson. Globally trained handwritten word recognizer using spatial representation, convolutional neural networks and hidden markov models. In *NIPS-6*, San Mateo CA, 1994. Morgan Kaufmann.

[3] J.R. Ward and T. Kuklinski. A model for variability effects in handprinting with implications for the design of handwriting character recognition systems. *IEEE Trans. Syst., Man, Cybern.*, 18:438–451, 1988.

[4] Y. Le Cun. Generalization and network design strategies. In R. Pfeifer, Z. Schreter, F. Fogelman, and L. Steels, editors, *Connectionism in Perspective*, Zurich, Switzerland, 1989. Elsevier.

[5] Y. Le Cun, O. Matan, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, and H . S. Baird. Handwritten zip code recognition with multilayer networks. In IAPR, editor, *Proc. of the International Conference on Pattern Recognition*, Atlantic City, 1990. IEEE. invited paper.

[6] C. C. Tappert, C. Y. Suen, and T. Wakahara. The state of the art in on-line handwriting recognition. *IEEE Trans. Patt. Anal. Machine Intell.*, 12(8):787–808, 1990.

[7] I. Guyon, P. Albrecht, Y. Le Cun, J. Denker, and W. Hubbard. Design of a neural network character recognizer for a touch terminal. *Pattern Recognition*, 24(2), 1991.

[8] S. Wu and U. Manber. Fast text searching with errors. Technical report, Department of Computer Science, University of Arizona, TR-91-11, 1991.

[9] R. Baeza-Yates and G. Gonnet. A new approach to text searching. In *Proceedings of the 12th Annual ACM-SIGIR Conference on Information retreival*, pages 168–175, Cambridge, 1989.