Optical Character Recognition for Self-Service Banking

Lawrence D. Jackel

David Sharman

Charles E. Stenard

B. Ivan Strom

Derek Zuckert

Self-service automatic teller machines (ATMs) have dramatically altered the ways in which customers interact with banks. ATMs provide the convenience of completing some banking transactions remotely and at any time. AT&T Global Information Solutions (GIS) is the world's leading provider of ATMs. These machines support such familiar services as cash withdrawals and balance inquiries. Further technological development has extended the utility and convenience of ATMs produced by GIS by facilitating check cashing and depositing, as well as direct bill payment, using an on-line system. These enhanced services, discussed in this paper, are made possible primarily through sophisticated optical character recognition (OCR) technology. Developed by an AT&T team that included GIS, AT&T Bell Laboratories Quality, Engineering, Software, and Technologies (QUEST), and AT&T Bell Laboratories Research, OCR technology was crucial to the development of these advanced ATMs.

Introduction

Automatic teller machines (ATMs) have provided additional convenience to everyday living. No longer must customers visit their banks during specific hours to withdraw cash or transfer funds between accounts. It is now possible for a customer whose bank is located in New Jersey, for example, to withdraw cash in California, Hawaii, or even Australia.

AT&T Global Information Solutions (GIS) is the leading supplier of self-service banking systems and has more than 40 percent of the worldwide market share (excluding Japan). This translates to an installed base of approximately 120,000 units located in more than 100 countries. In 1994 alone, GIS shipped about 32,000 units—a record year for sales.

Product innovation is one way in which GIS intends to maintain and increase its market leadership. Cutting-edge advancements in *optical character recognition* (OCR) from AT&T Bell Laboratories is providing the technology needed for such innovation.

Although enormously successful, ATMs typically can provide only two of the multiple services a human teller provides—the dispensing of cash and transferal of funds accompanied by a printed acknowledgment. Another teller service—and a major self-service opportunity—is the acceptance of payments or deposits that normally involves handling some form of paper documentation.

Until now, ATMs could not perform transactions based on paper input, except when a document was placed in an envelope for subsequent manual processing. Examples of such paper transactions included check deposit, check cashing, and bill payment.

Recently, GIS released a new generation of ATMs that can perform paper-initiated transactions, thus providing more comprehensive customer service. One of these new machines, the AT&T 5665 is shown in Figure 1.

ATMs must "read" either machine-generated or hand-written characters to complete such paper transactions. OCR makes possible the process of recognizing inked characters on paper.

Commercial OCR systems for reading

Figure 1. Pictured is the new AT&T 5665 OCR-enabled automatic teller machine (ATM). This ATM, which allows customers to process checks and pay bills 24 hours a day, features advanced character recognition technology. This latest generation of ATMs is the result of a partnership between AT&T Bell Laboratories Research; Quality, Engineering, Software, and Technology (QUEST); and AT&T Global Information Solutions.

"clean," machine- and hand-printed materials have become available in the last few years. Both reliability and accuracy, however, severely degrade or the systems even fail on low-quality character images—either machine generated or hand written—that appear on checks. Thus, new character recognition methods were needed to obtain acceptable performance.

Panel 1. Abbreviations, Acronyms, and Terms

ATM—automatic teller machine
CAV—courtesy amount verification
giro—a financial instrument, similar to a check, commonly used in some European countries
GIS—the Global Information Solutions business unit of AT&T
LeNet—a patented, highly accurate family of neural-network character recognizers developed by AT&T Bell Laboratories
OCR—optical character recognition
QUEST—the Quality, Engineering, Software, and Technologies organization of AT&T Bell Laboratories

Overview of OCR Technology

The difficulty in performing accurate character recognition depends greatly on the quality and context of the characters to be read. It is relatively easy to attain high accuracy when reading clean, machine-printed documents. This is particularly true if only one font is used and if it is known in advance—or from the context—whether digits or letters are being read. Printed text is more difficult to read when the characters are scattered throughout a document that contains elements of noncharacter images, such as decorative line art or photographs.

Similarly, hand-written recognition is easiest when the characters are placed on forms in well-defined fields, with each character occupying its own box. If either digits and letters or upper- and lower-case letters are mixed, additional information is still needed to avoid character ambiguity. For example, context is needed to distinguish the number "0" (zero) from the capital letter "O" (oh), or a lower case "l" (el) and a capital "I" (eye) from the number "1" (one).

Particularly challenging recognition tasks appear when attempting to read the dollar amounts written or printed on checks. Character recognition on personal checks is most difficult. No constraints exist on *how* the characters must be written, although their general locations are known (Figure 2). Hand-written characters often touch each other, too, making it difficult to tell where one character ends and another starts. Furthermore, breaks in the ink pattern often represent part of a printed character so that several ink blobs must

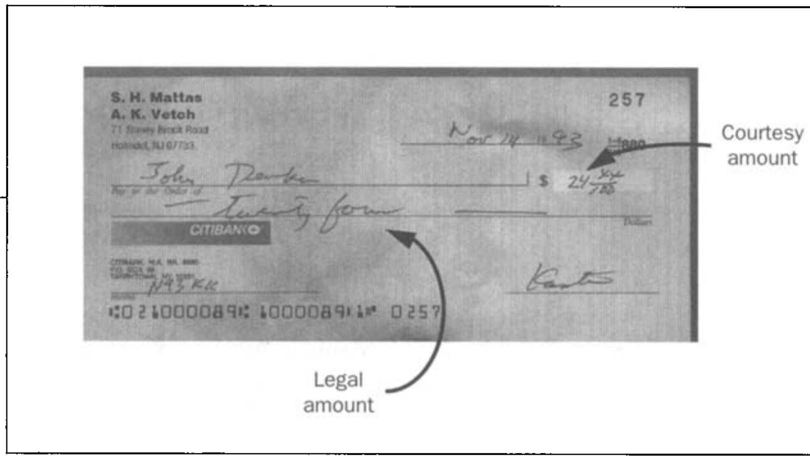

Figure 2. Optical character recognition technology presents some interesting challenges when attempting to read the dollar amounts written or printed on checks. Character recognition on personal checks is most difficult. No constraints exist on *how* the characters must be written, although the general areas in which they appear are known.

be combined to create one complete character.

Different problems must be solved when reading United States business checks. Unlike personal checks, the U. S. banking industry has not established strict layout guidelines. For example, the dollar amount can be written almost anywhere on a check's face.

The following sections describe the *recognition engine*, which is used to identify ink blobs as particular characters. The concept of *courtesy amount verification (CAV)* is discussed next, which couples user-supplied information with character recognition, allowing the attainment of acceptable system performance. The paper then progresses to the section on the *CAV check-reading process*, which covers image capture through the acceptance or rejection of a transaction. Before concluding, the paper discusses some additional services made possible through advanced ATMs.

The Recognition Engine

The combination of high intrinsic accuracy and easy customization for the particular region in which the ATM will be used is an exclusive feature of the recognition algorithms resident in AT&T OCR-enabled equipment. This feature results from the methodology used to "train" the recognition engine and its architecture.

The recognizer is a type of *neural network* customized for character recognition. A neural network is a computer program that performs nonlinear transformations of input vectors. For the specific task of pattern recognition, the network transforms the input bitmap image into a list of probabilities of the classes of characters that might be represented by the image—for example, certain digits or letters. The nature of the transformation is determined by the architecture of the network and the particular value of a set of network parameters, which are known as *weights*.

An important feature of neural networks is that they can be trained from examples to perform the desired recognition task. Such training is accomplished using the procedure shown in Figure 3. To train a network, a database of sample characters is needed. Each sample must have the correct label of the character class, which is known as the *target*.

Initially, the weights in the network are set to random values. Character images are then presented to the network, which produces a list of numbers (ten numbers in the case of digits) that represents the likelihood that a character belongs to a particular class. After each presentation—and by using a mathematical process known as *gradient descent*—the weights are adjusted to make the network output closer to the target. By repeating this process using a large training database, a set of weights can be determined. These weights not only will provide accurate output classification scores for the training examples. Most importantly, they will accurately classify new characters not contained in the training database.

This ability to classify previously unseen characters is crucial to ATM applications. As long as training examples exist that are similar to the characters appearing on users' checks, the recognizer should make the correct classification. For this reason, a large training database is needed—one that contains the characters people

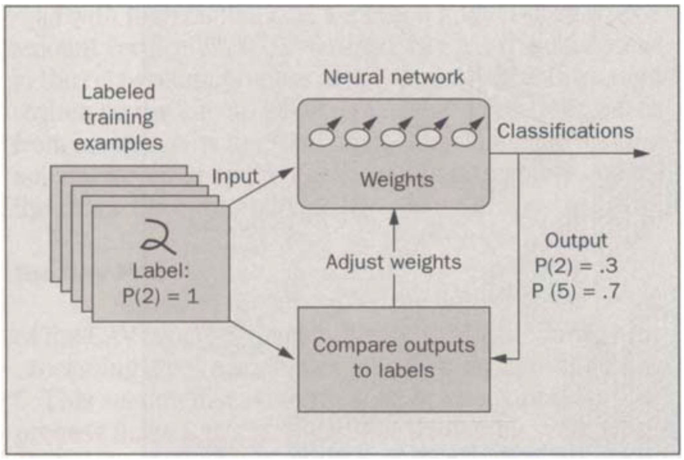

Figure 3. This diagram illustrates the training procedure used for the neural network. The network has a set of adjustable parameters called *weights* that are set by the training procedure. Training images are presented to the network. The network's output is compared to a training example's label. The weights are then adjusted to bring the output closer to the label. This process is repeated for thousands of examples. If the network is designed properly, it will eventually "learn" to classify correctly nearly all the training examples. More importantly, it will attain high accuracy classification on new test images that are not part of the training set.

normally write in the geographic regions in which the ATMs will be deployed.

Neural Network Architecture. In addition to having a large and representative training set, another key to good recognizer performance is the *neural network architecture*. In general, neural networks consist of layers of computational units in which information from one layer is fed to the following layers through the weights discussed earlier.

Each layer contains *units*. The number of units in the input layer corresponds to the number of pixels in the input image. The number of units in the output layer corresponds to the number of output classes—ten units, in the case of a digit recognizer. The number of units in the intermediate *hidden layers*, the number of hidden layers, and the information flow from one layer to the next is determined by the network designer. It is important to note that great care must be taken in the design of the network to achieve high accuracy.

Working together, staff from the AT&T Bell Laboratories Quality, Engineering, Software, and Technologies (QUEST) and Research organizations designed the highly accurate *LeNet* family of patented, neural-network character recognizers. Yann LeCun of AT&T Bell Laboratories in Holmdel, New Jersey, designed the first LeNet recognizer.[1] He also contributed to most of the later versions.

Figure 4 shows the typical architecture of a LeNet recognizer. These recognizers are structured so that alternate layers in the network perform two-dimensional feature extractions. The first layer of feature extraction detects simple objects in the image, such as edges. Subsequent layers detect more complex structures. Other layers, located between the feature extraction layers, perform nonlinear smoothing and subsampling. These layers make the network less sensitive to image distortions.

A distinguishing characteristic of LeNet recognizers when compared with other recognizers is that the features extracted are determined through weight adjustments during the training process. Thus, the features are optimized automatically for the task of recognizing characters, like those found in the training set.

Recognition accuracy depends both on the particular aspects of the network architecture and the number of training examples. The more training examples, the higher the accuracy. However, for a fixed architecture (a fixed number of feature maps and layers), the achievable accuracy is limited no matter how large the training set. Additional feature maps and/or layers must be added to the network to take full advantage of a large training set.

It is also important that the network should not be too complex for the amount of training data available. In trying to use a complex network that has many weights (for example, 100,000) with a small training set (for example, 10,000 examples), poor performance would result. The reason for this is there is insufficient training information available to set the weights properly. More accurate classifications could have been obtained using a simpler network with fewer weights.

In general—for highest accuracy—the number of weights should be comparable to the number of training examples. When network complexity is properly matched to the number of training examples, the error

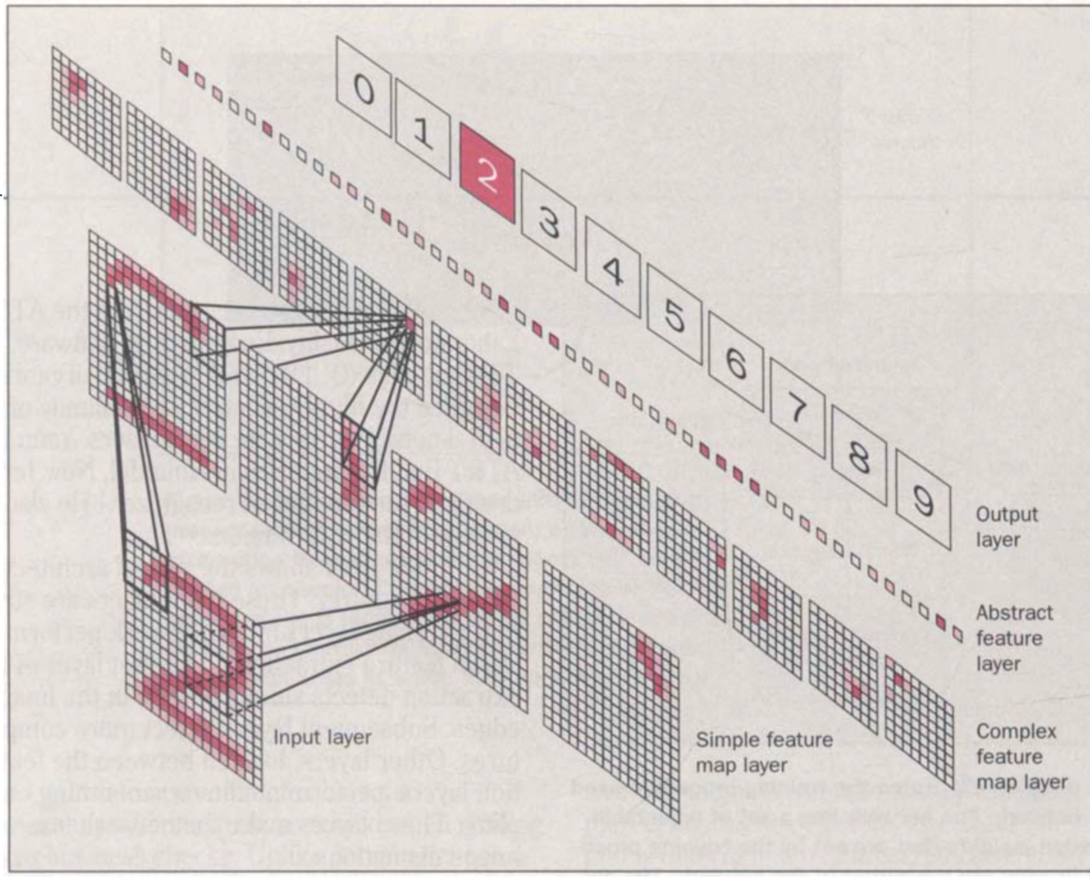

Figure 4. This drawing represents the typical architecture of a LeNet recognizer, which has a series of layers that perform feature extraction. A unique strength of LeNet—distinguishing it from other character recognizers—is that the particular features extracted are determined through the weight adjustment of the training process. Thus, the features are automatically optimized for the task of recognizing characters like those in the training set. Located between the feature extraction layers are other layers (not shown) that perform nonlinear smoothing and image subsampling, reducing the number of pixels in the feature maps. The subsampling desensitizes the recognizer to small shifts, rotations, and distortions of the input image.

component on test data decreases by about a factor of two for a tenfold increase in training-set size. On standard data sets supplied by the U. S. National Institute of Standards and Technology, the best AT&T recognizers achieve test accuracy greater than 99 percent on isolated hand-written digits when trained with 60,000 examples.[2]

The notion of properly adjusting the complexity of a network to match the size of a training set is part of a useful and elegant theoretical framework for pattern recognition. This framework is known as *structural risk minimization*. Its principal inventor, Vladimir N. Vapnik of AT&T Bell Laboratories in Holmdel, New Jersey, has written several landmark books on the subject.[3] Recent work at AT&T Bell Laboratories, based on Vapnik's concepts, facilitates the prediction of how the accuracy of a recognizer will improve as the size of the training set is increased.[4] This information is used to determine the quantity of training data required to meet the needs of GIS customers.

Courtesy Amount Verification (CAV)

If the best OCR systems are used to read the courtesy amount on personal checks, high-confidence recognition (an error probability of less than one percent) will be achieved only on about 50 percent of checks. While this is an acceptable read rate for applications in which a reduction of keypunching by human operators is desired, it is far too low to satisfy user demands for check cashing.

To increase the number of checks that can be read with high confidence, a method known as courtesy amount verification (CAV) is used. In CAV, the user assists in the recognition process simply by keying in the amount written on the check. Instead of having to read the check from scratch, now the ATM must determine only that the amount keyed in by the user is a high-probability match to the characters written in ink.

The CAV Process

Although the character recognizer is the heart of the CAV process, many other sophisticated image-processing steps are required for CAV, as shown in Figure 5. This section discusses the CAV or check-reading process that occurs after an optical imaging system creates a bitmap image of an ATM user's check. The principal developers of the CAV system are Jan Ben; Jane Bromley; Chris Burges; Janet Lee; Madjid Mousavi; and Craig Nohl, all members of the AT&T Bell Laboratories QUEST organization; and Mike Stanton (resident visitor).

Courtesy Amount Location. In the United States, industry standards for personal checks require that the courtesy amount field be located in a particular region (Figure 5a, b). This standard is used to locate the region that contains the courtesy amount. The courtesy amount can be found by locating the required, printed dollar-sign symbol. The courtesy amount is situated to the right of the symbol, reducing the search region's size. This step is shown in Figure 5c.

Image Cleanup. Some checks are decorated with artwork that can confuse a character recognizer. Even though LeNet generally ignores extraneous ink blobs in the input image field, recognition accuracy is higher if the blobs not belonging to specific characters are removed. The problem, of course, is removing only the superfluous blobs while leaving intact those legitimately belonging to characters.

Such clean-up can be done by passing geometric filters over the image and creating feature maps similar to the first-level maps produced by LeNet. In this case, the extracted features are line shapes and thickness that span the range found in courtesy amount characters. Clean images of the courtesy amount field can be generated, as shown in Figure 5d, by combining the feature maps and using the knowledge that characters must fall within certain ranges of overall size and shape.

Locating the Dollars and Cents Fields. Now that clean images represent the courtesy amount, the characters that differentiate cents from dollars can be identified. In the United States, checks are written in such a way that the cents field is always to the right of the dollars field. Several methods of representing cents, however, can be used. These include fractions, decimal points, lines, and characters like "xx."

After examining many checks, a simple "grammar" was developed that describes the ways in which different individuals write the cents field. The grammar is used to identify the cents field and separate it from the dollars field. Therefore, the amount in each field is recognized individually. For dollars recognition, only the digits in the dollars field are recognized. For cents recognition, symbols like "xx," "00," and "/" must first be removed to avoid confusion.

Simultaneous Recognition and Segmentation. When attempting to recognize a string of characters like those in the dollars field, it is often difficult to determine which ink blobs belong to a particular character. The task of assigning blobs to certain characters is known as *segmentation*. Segmentation is easy if no characters are composed of more than one ink blob and no characters are touching. In this case, a *connected components analysis* is used in the recognition process.

In check images, no such guarantee exists. One must assume that characters might be touching or broken. This situation seems to create a dilemma—characters cannot be recognized until they are first segmented. On the other hand, segmentation cannot be completed unless the characters are known in the first place. This dilemma is resolved by segmenting and recognizing simultaneously. (Segmenting accurately is difficult. Thus, forms designed for the machine recognition of hand-written characters require that the characters be written in boxes, eliminating the need for segmentation.)

To start the recognition and segmentation process, ink blobs are found that—when properly combined—create complete, isolated characters. The blobs used are all the isolated blobs in the field augmented by pieces from larger blobs, which were cut using simple rules designed to break apart blobs that include ink from more than one character.

Next, ink blobs are taken either singly or in groups to create "legal" string segmentations (Figure 5f). Legal segmentations are those that use all the ink once

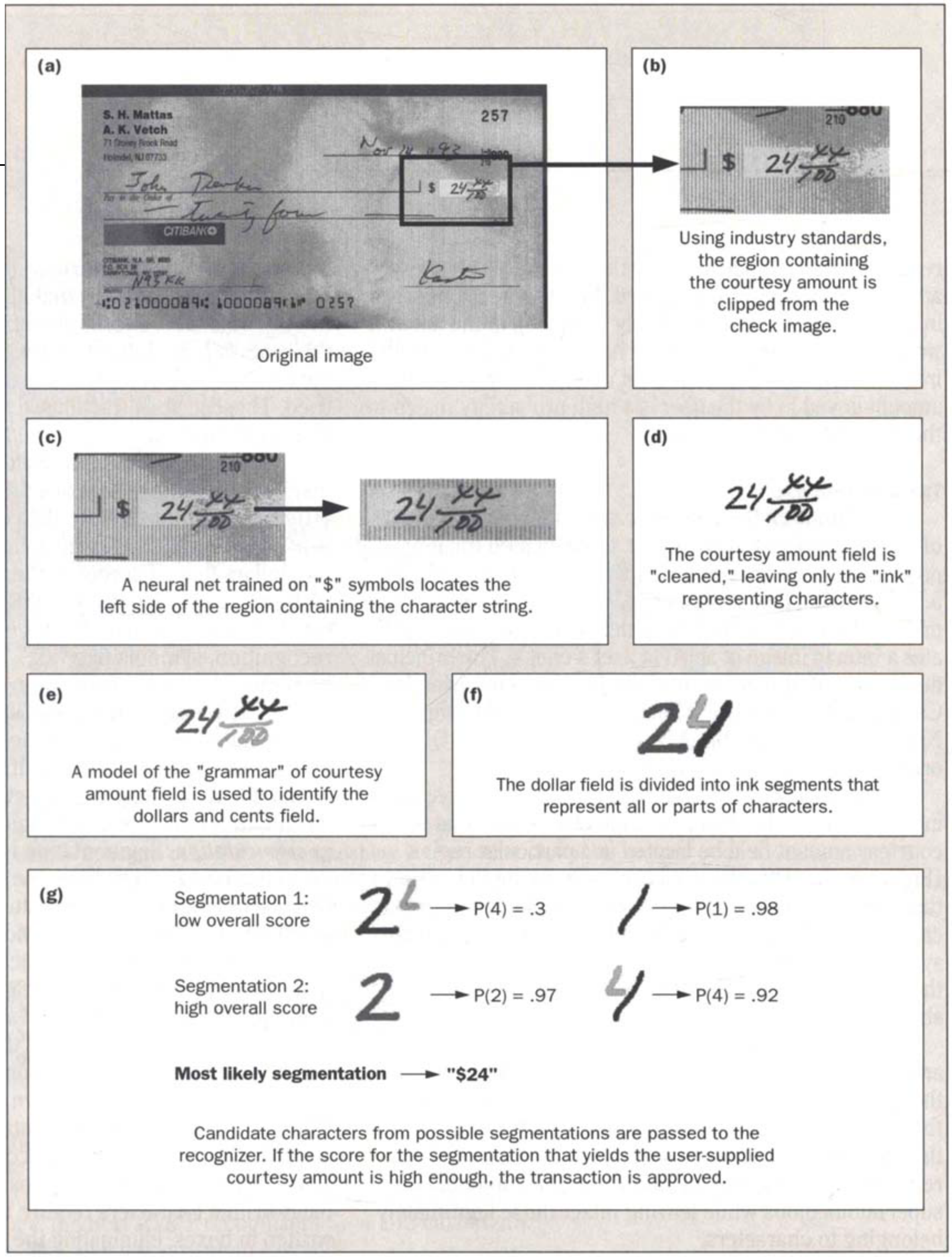

Figure 5. This illustration shows the seven-step process typically used in verifying the courtesy amount on a check.

but only once. At this point, the task becomes easier because constraints from the CAV process limit the legal segmentations. In particular, the ATM user has already entered the amount believed to be written on the check. Therefore, the total number of characters in the string is known. In the example shown in Figure 5, a user will have keyed in "$24.00," so exactly two characters must be sought in the dollars field.

Blobs that can represent characters as part of a legal segmentation are known as *candidate characters*. All the candidate characters are presented to the recognizer and the recognition scores recorded. An efficient graph-traversal technique is then used to determine if the courtesy amount entered by a user corresponds to a high-confidence string recognition. If this is the case, the transaction can proceed, as shown in Figure 5g.

If a good match is not obtained, several options are available, depending on bank policy. The bank can ask the user if a mistake was made. If so, the user can re-enter the courtesy amount. If the user re-enters the same amount, the bank can either reject or approve the transaction, but the bank informs the user that the transaction will be verified for accuracy later.

Presumably, banks would cash only those checks that do not exceed the verified balances in users' accounts. The strictness of a bank's policy depends on how reliably the recognizer gives correct strings high scores and incorrect strings low scores. Typically, a bank will want to know the number of correct transactions accepted while ensuring that incorrect acceptances are maintained below a predetermined percentage. With a perfect recognizer, all correct transactions are accepted and all incorrect transactions rejected. In the current systems GIS is deploying, recognition is sufficiently accurate that nearly all correct transactions are accepted and most incorrect ones rejected. Thus, few legitimate users are disappointed.

The next big jump in ATM performance will occur when the courtesy amount is cross-checked against the legal amount (Figure 2). The algorithms for legal amount verification are being developed and will be available in 1996 for integration with the CAV solution.

Additional Services

CAV enables an ATM to cash checks and confirm deposits. Having accurate OCR in ATMs, however, facilitates additional customer services. The specific services provided depend on local practices in the country in which the ATM is installed. For example, customers can pay their utility bills at banks in many locations in the United States. Using a procedure similar to CAV, bills can also be paid at OCR-enabled ATMs.

In some European countries, an instrument known as a *giro* is used. Like a check, a giro transfers funds from one account to another and is mainly used to pay bills. Typically, a merchant sends a giro or payment coupon to a customer. The giro is imprinted with the customer's name, address, and bank account number. It also contains the merchant's bank account number, the amount due, and due date. The customer signs the giro, authorizing payment (that is, the transfer of money from the customer's account to the merchant's account) and submits it to the bank. Using the new ATMs, however, this procedure is streamlined, allowing customers to process giros conveniently at any time.

In automating giro processing, the ATM's OCR equipment reads the relevant data from the giro—the "to" and "from" account numbers (customers may have several accounts), amount to be paid, and due date. This capability saves the time and effort required to enter the information by hand. Similarly, it reduces the bank's administrative and processing costs.

The following anecdote underscores the convenience a self-service ATM can provide. Recently, GIS field engineers installed an AT&T 5665 OCR-enabled ATM at a customer's site in Helsinki, Finland. Unwittingly, the engineers parked their vehicle in a restricted zone and returned later to find they had been issued a summons. Realizing the convenience of the ATM they had just installed, the engineers immediately returned to the machine, inserted the ticket, and paid the required fine.

Conclusion

This OCR-enabled ATM project is a good example of how rapid product realization results in new services becoming available to AT&T customers worldwide. The project was successfully completed by a closely knit AT&T product-development team able to envision the possibilities, focus the technology, and anticipate the market opportunities for enhanced ATM services. Thus, AT&T was the first to make these services available, and this timely entree helped increase AT&T's market share in the highly competitive ATM arena.

Acknowledgments

The authors wish to thank the many AT&T contributors who collaborated to make OCR-enabled ATMs a reality. The team included individuals in the Engineering Group from AT&T-GIS, St. Davids, Scotland; the Advanced Information Systems Development and the Advanced Information System Engineering groups in QUEST; the AT&T-GIS Human Interface Technology Center in Atlanta, Georgia; and the AT&T Bell Laboratories Adaptive Systems Research Department in Holmdel, New Jersey. Accurate training of the neural networks was aided by the use of the character distortion model developed by Henry Baird at AT&T Bell Laboratories, Murray Hill, New Jersey.

References

1. Y. LeCun, O. Matan, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, and H. S. Baird, "Hand-Written Zip Code Recognition With Multilayer Networks," *Proceedings of the International Conference on Pattern Recognition*, Atlantic City, New Jersey, 1990.
2. L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, L. D. Jackel, Y. LeCun, U. Muller, E. Sackinger, P. Simard, and V. N. Vapnik, "Comparison of Classifier Methods: A Case Study in Hand-Written Digit Recognition," *Proceedings of the International Conference on Pattern Recognition*, Jerusalem, Israel, 1994.
3. See, for example, V. N. Vapnik, *Estimation of Dependences Based on Empirical Data Dependences*, Springer-Verlag New York, Inc., 1982.
4. C. Cortes, L. D. Jackel, S. A. Solla, V. N. Vapnik, and J. S. Denker, "Learning Curves: Asymptotic Values and Rate of Convergence," *Advances in Neural Information Processing Systems*, Vol. 6, edited by J. D. Cowan, G. Tesauro, and J. Alspector, Morgan Kaufman Pubs., Inc., San Francisco, California, 1994.

(Manuscript approved June 1995)

Lawrence D. Jackel *is head of the Adaptive Systems Research Department at AT&T Bell Laboratories in Holmdel, New Jersey. His department is responsible for developing automatic computer learning tools that can be applied to tasks in pattern recognition and data mining. Mr. Jackel, who joined AT&T Bell Laboratories in 1975, has a Ph.D. degree in physics from Cornell University, Ithaca, New York.*

David Sharman *was team leader for the AT&T 5665 ATM during its development. He is now coach of the product group at AT&T Global Information Solutions in St. Davids, Scotland, with responsibility for all ATM products. Mr. Sharman joined AT&T in 1990, and he has a Ph.D. degree in image processing from Strathclyde University, Glasgow, United Kingdom.*

Charles E. Stenard *is technical manager of the Advanced Information Systems Engineering Group at AT&T Bell Laboratories in Holmdel, New Jersey. His group is responsible for research transfer and technology development relating to image text understanding. Mr. Stenard has a Ph.D. degree in mathematics from Princeton University in New Jersey. He joined AT&T Bell Laboratories in 1967.*

B. Ivan Strom *is technical manager of the Advanced Information Systems Development Group at AT&T Bell Laboratories in Holmdel, New Jersey. His group is responsible for developing the optical character recognition hardware and software subsystems for use in ATMs and other AT&T products and services. Mr. Strom joined AT&T Bell Laboratories in 1977, and he has a Ph.D. degree in computer science from Columbia University, New York City.*

Derek Zuckert *is the team leader for recognition software at AT&T Global Information Solutions in St. Davids, Scotland. He is responsible for defining and implementing the self-service recognition solution strategy. Mr. Zuckert, who joined AT&T in 1990, has a B.S. degree with honors in electrical and electronic engineering from the Robert Gordon Institute of Technology in Aberdeen, United Kingdom.*