The recognition of handwriting

How a successful model of vision can be obtained by algorithmic information processing

J.-C. Simon†
A2iA, Paris France

Abstract. Only 15 years ago, the recognition of handwriting was considered as Science Fiction. Today ICR[1] Softwares executed on simple PCs allow to identify the amount of 70% of handwritten checks with the same rate of confusion that human operators, this not only on figures but also on words. Today 60 sites in France, UK, USA are using these process in production. But to reach this state several difficulties had to be solved, a company had to be started to finance research and application. The scientific interest of this endeavor is several: Show how a model of vision could be obtained by algorithmic information processing; introduce experimental method in a formerly philosophical interpretation; set up new scientific concepts; show that algorithmic models may not be only mere applications of mathematical models.

Résumé. Il y a 15 ans seulement la Reconnaissance Automatique de l'écriture manuscrite se limitait à des essais de laboratoire, à la limite de la Science Fiction. Notre recherche a permis de développer un logiciel qui est aujourd'hui installé dans 60 sites industriels de production et reconnaît journellement les montants de 20% des chèques manuscrits français, non seulement en chiffres mais aussi en lettres. Il a fallu résoudre d'innombrables problèmes scientifiques et techniques, et pour cela développer dans des conditions difficiles une entreprise en France et à l'Etranger. L'intérêt scientifique de cette étude est multiple : dégager des concepts scientifiques nouveaux; modéliser un domaine jusque là purement spéculatif; introduire la méthode expérimentale; montrer la validité d'une modélisation informatique autre qu'une simple mise en œuvre d'un modèle mathématique.

Keywords: handwriting recognition, information processing

Mathematics Subject Classifications (2000): 68T35, G8T10, 68Q30

A new type of models in science

Since Galileo, the generally agreed paradigm is that Mathematics will model the World. The success of this principle has been the success of Science, in so many domains, Mechanics, Astronomy, Relativity, Quantum Physics. The predictive power of the mathematical models is sometimes astounding as it is witnessed in Particle Physics. The modern powerful computers allow to apply the mathematical algebraic or differential equations, i.e., to compute solutions. But this approach does not seem productive for information processing, and many new domains are indeed relevant to information processing: Economy, Biology in particular Molecular Biology, and all of the brain processes such as vision, hearing, language, thinking. Certainly a knowledge of mathematics is useful, but no one has ever found the equations that may explain brain processing.

[1] ICR is an acronym for Intelligent Character Recognition.

G. Dito and D. Sternheimer (eds.), Conférence Moshé Flato 1999, Vol. 1, 73 – 80.

Our study of handwriting recognition demonstrates that new type of models may be build up without the Platonic approach: Conceive a mathematical model, then compute solutions. We call **models of 2nd type** these new algorithmic models, implemented directly by programs.Then “the program is the model”.

This new approach does not imply that any hacker will program successfully such 2^{nd} type models, by cut and try. One has to be guided by concepts, ideas and experiments. An important scientific community deals with the Pattern Recognition – **PR** – problems. IAPR, International Association for Pattern Recognition, is an umbrella association which covers 38 national PR associations, the members of which are in Universities, Engineering Schools, Industries. Applications are coming on the market in speech recognition, robotics, document understanding and recognition (cf. [1, 2, 3, 4, 5]). Let us show what are the main scientific concepts which are our guidelines in handwriting recognition.

Important concepts in handwriting recognition

Principles common to all algorithmic Pattern Recognitions

Basically PR goal is to interpret data. Let us limit to images. An image data is obtained through an optical sensor, which interprets the outside world, the result of which is a set of pixels[2], which may have several values, or only two, then they are said binary. To be processed by computers, a pixel is represented by a number of bits, one if it is a binary pixel. The image of a A4 page has currently $n = 3.6$ million pixels. The number of such possible binary images is 2^n, an unthinkable number. One could describe such an image as a point in a space of 2^n dimensions. The PR problem is to obtain projections of this point in sets of small cardinality, for example the set of names of the objects in the image. It is clear that this cannot be done in one step.

A chicken and egg problem: Segmentation and recognition

To find the sub images of interest, i.e., representing an object, we have to identify or at least determine the class of the object. But for this we have to segment into sub images. Clearly a vicious circle. The way out is to find simple *primitive features*, and as we see later regular features.

Finding primitive features

Primitive features in a digital image are textures, edges, lines, loops, any elementary features, which are found by the corresponding *operators*. An operator is an algorithm implemented in a program executed by a computer. One may consider it as a black box, the input data of which is a sub image, the output being a “soft decision”: A score of credibility – similar to a probability – and parameters of the feature. In fact it is not possible to separate feature and operator: *An operator is the constructive definition of a feature*, whether or not it corresponds to what a human observer estimates the feature to be.

2 A pixel is the result of the measure of the illumination amplitude in a small solid angle.

Regular versus singular features
Experience has shown that some features are more easy to find. Taking the example of speech, 95% of vowels are correctly found, instead of 40% of consonants. The vowel signals are (pseudo) periodic, a consonant being an accident of a vowel. The operators finding the vowels are *invariant to translation*. This is true also for texture detection in images. We give the attribute of *regular* to periodic features and to the corresponding operators invariant to translation. Features which are not regular are *singular*. The interest is to find first the easy regular features. What is left is singular features, more easy to detect as they are then segmented. This approach has been used successfully in handwritten word recognition.

Operator complexity
The price to pay for constructive process is computing complexity. The execution of an algorithm requires a number of elementary steps function of the size n of the input data. When n is large enough, the complexity is measured by an order of magnitude $O(f(n))$. Most of the PR operators imply an exponential complexity $O(e^{\alpha n})$. As soon as n is large, though the number of steps is finite, it is considered infinite in practice, and not practicable no matter what the computing power is. Thus, operators of lower complexity have to be used.

Exchanging precision for speed
Very much faster polynomial operators of $O(n^{\alpha})$ are in use, preferably linear. But their precision is lower; in other words they introduce errors. The price to pay to come out of the time deadlock is to lower precision. At this point, we may remark that regular operators are of low complexity, which facilitate the approach suggested above.

Combination of operators
Experience has shown that by appropriate combination of several operator outputs, precision could be restored. Many research efforts were done to find best combinations. It is now admitted that a linear combination of the logarithms of the credibility scores using a Neural Net learning is close to the optimum. It was found also that such combinations of several low complexity operators reduce the generally *chaotic character of algorithms*, i.e., a small change in the input data would cause a large change in the output results.

Multiple levels of decision
The recognition process is performed in several levels of *interpretation*. At each, operators (algorithmic programs) give an *interpretation* of the input *representation*. The first level is the optical sensor level which interpret the outside world as a digital image. This data is the representation to be interpreted as features at the next level. These features will be the representation of the next level being interpreted as objects; and so on.

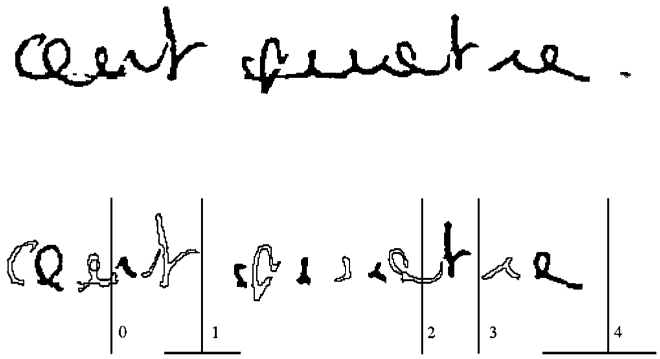

Figure 1. Words segmented in graphemes.

Soft decisions
Instead of "hard" decision by Yes or No, the result of an operator or of a combination of operators is a "soft " decision, i.e., a *list* of possible features or objects each with a credibility score. The next recognition level uses the set of lists to set up a new set of lists of possible objects with credibility scores

Increased abstraction
At each level abstraction is increased, and localization is decreased. For example if a letter 'e' is found in a sub image of $p \times p$ pixels, the information of the letter does not give the precise sub image. A huge number of sub images $p \times p$ can represent an 'e'. What is gained in abstraction is lost in localization (extension).

Handwriting recognition
Using letters to recognize a words
Figure 1 shows on the top the binary image two French words "cent quatre". Using the approach to find first the regular features, the words can be segmented in singular features, the *graphemes*, which are shown alternatively in black and white. A letter is made of one grapheme – like the 'c', the 'e' – or two graphemes – like the 'n', the 'u', or more like an 'm'. So the problem of segmentation in letters is well advanced. On Figure 1, the possible segmentations in words is also shown by vertical bars. The correct ones having an horizontal bar. But we do not know them a priori. So all combinations of segmentation have to be examined. Inside of each sub image the likelihood of a word is examined. For this all combination of one or two graphemes generates a possible list of couples (letter, score). Finding the optimal paths among these list generate a new list of couples (word, score). And so on up to a list of phrases- in the considered application a check amount. At each step, except the grapheme generation, all decisions are "soft". To have to consider many segmentations may generate a "combinatorial explosion" and again an impractibility. The credibility scores are of course essential to keep down the number of cases considered.

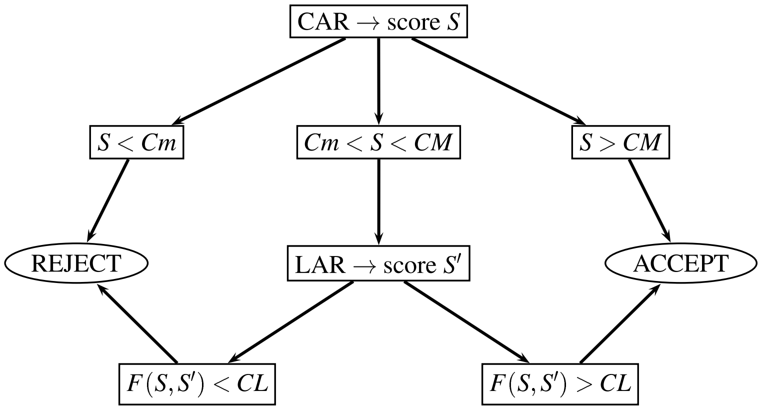

Figure 2. The use of CAR and LAR.

A global recognition
Another approach is to identify a word without finding the letters. Then the features are ascenders, descenders, loops, horizontal bars, etc.

Application to check amount recognition

5 billion checks in France, 60 billion in the USA are issued each year. Their amounts have to be coded to be processed in payment systems. Up to now it is done "by hand". It is a costly and unrewarding activity, which should be automatized. The amount of a check is coded twice, by numbers – the "courtesy amount" –, by words – the "legal amount". **CAR** is the acronym for Courtesy Amount Recognition, **LAR** for Legal Amount Recognition. The ICR Software "A2iA Check Reader" recognize both CAR and LAR, according to the scheme of Figure 2. CAR delivers a score S. If $S < Cm$ the image is rejected. If $S > CM$, the amount is accepted. Cm and CM are two thresholds which may be adjusted. In the grey zone, $Cm < S < CM$, LAR is performed giving a score S'. Let CL be another threshold; if $S' < CL$ the check is rejected. Otherwise the two CAR and LAR lists of couples – amount and score – are combined to give a final amount with a score, which is accepted of rejected according to a third threshold CL.

Applications in production
At the beginning of Year 2000, sixty payment system sites are using "A2iA Check Reader", mostly in France, but already also in the USA and UK. These ICR softwares are reading now of the order of five million check/day. Data entry – reading a check amount – by software is done by an ordinary Pentium 400MHz PC under Windows NT, at the speed of two amounts per second, 7000 per hour. If a higher speed is necessary, images are distributed among N PCs; the speed is multiplied by N. The ICR Software itself, written in 'C', has 85 000 lines of code. It is constantly improved and has also other added possibilities. A good human operator is coding

approximately 2 000 check amounts per hour, with a substitution error of 1%. By adjusting the thresholds *Cm*, *CM*, *CL*, the substitution rate of the ICR software can be kept down statistically at 1%. Then the ICR acceptance rate is in the range 65% to 75%, depending of the type of check, the quality of the image. The 35% to 25% rejected checks are read by human operators, who still beats the machine on difficult checks.

Concluding remarks

The A2iA team was quite lucky to be able to bring a basic research to the state of a massive application in the banking industry. Other applications will follow in the Postal address reading and more generally in document processing and recognition. This allows us to carry on research in the important field of *information processing and recognition*. Certainly we are not alone: An important scientific research community [6] is working on these topics, and many high tech companies are created.

Is information processing a science?
If Information Processing (I.P.) does not follow the Galilean paradigm: "Mathematical Models will explain the World ", – eventually may allow constructive predictions –, and if this is the necessary condition of a science, then I.P. is *not a science*. On the other hand, if I.P. gives constructive models of natural phenomena, which cannot be explained, modeled, or predicted by Mathematics, it should be *deemed to be a Science*. In fact a new scientific paradigm, to which we give the name of Models of 2^{nd} type. It is easy to see that these models will predominate in fields where Information Processing, more precisely representations and interpretations, are basic: Economy, Molecular Biology, and all of the brain processes such as vision, hearing, language, thinking. Then the performances will be the proof of validity of the concepts which guide the build up of such 2^{nd} type models.

Brain processes
The Nobel laureate Francis Crick wrote in *Scientific American 79*: "If most of philosophical thought is barren since 2000 years, and will likely remain so, it is because philosophers did not learn to understand the language of information processing". Brain processing, vision, language, understanding have been mostly a philosophical field, and *we know very little of the brain processing as an information processing system*. Knowing a lot about the neurons, does not help much to understand how 100 billion neurons are achieving the extraordinary brain processes, seeing, hearing, understanding, thinking, speaking.Building up constructive models of a brain process, even partial, is like introducing the experimental method in these domains, which are in a state similar to Medicine 200 years ago, when discourse was more important than practice.

Interest of vision models
A third of human cortex is devoted to vision, which is our most important sense. Vision is a lower level process in the sense that we have no control on our interpretation of visual perception: We cannot help to recognize objects. Our visual system is an *automatism*, which is not under our control, as are higher level processes like muscle control or thinking. Primitive animals like insects are visual automata. Vision appeared very early in the evolution of Life. It is likely that the brain processes of vision, which now are so refined in human beings, were the basis for most of the upper processes. Physiologists have found that a cortical zone, used for vision in the macaque monkey, has been confiscated for speech in human beings. Logical thinking appeared very recently in human beings. Instead of explaining brain process by logic, which is the strategy of Artificial Intelligence, it seems more appropriate to follow natural evolution and to try first to understand vision process. Though recognition of handwriting is a very small part of human vision, the algorithmic model, which has been developped, has similarities with the discoveries of brain physiologists: Neurons finding elementary features such as segments of different directions; layers of neurons to be compared to the several levels of recognition in our model; the many connections of neurons of a layer, to neurons of the next layer, similar to the model multiple combination of operators; at each layer, abstraction increases, etc. But a computer is not a brain. Our model uses the the most general and most efficient information processing machine available today to simulate vision. Our research has shown that the computer should not be used as a logical machine, but more and more as an analogue machine, to obtain good results. Thirty years ago, analogue computers were competitors of digital computers and could be started again with different technologies. But one point remains clear: We have to try to understand the brain processes to reproduce them, and use PR techniques rather that try to force some mathematical model. On the other hand, physiologists should be interested by constructive models. But more than ever, scientific interdisciplinarity seems more difficult to achieve, though many official declarations put it in the forefront of their preoccupations.

References

1. Duda, R. O. and Hart, P.: *Pattern Classification and Scene Analysis*, Wiley Interscience 1974.
2. Fukunaga, K.: *Pattern Recognition*, Academic Press, 1972.
3. Simon, J.-C.: *Patterns and Operators*, North Oxford Academic, 1986.
4. Bishop, C. M.: *Neural Network for Pattern Recognition*, Oxford University Press, 1995
5. Proceedings of ICPR, IEEE documents of 13 IAPR International Symposium,taking place every two years.
6. ICDAR '99, International Document Analysis and Recognition had 300 participants of 23 different nationalities, in September 99 at Bangalore India.

EDITORS' NOTE

Jean-Claude Simon, Professor Emeritus at Université Pierre et Marie Curie (Paris VI), Chairman and CEO of A2iA, died suddenly in Paris on February 16, 2000.

Jean-Claude Simon was brought to Paris VI from the private sector and always had an original approach to computer science problems. He has been a member of the Scientific Council of UAP, from its creation in 1984 by Yvette Chassagne and Moshé Flato, to its dissolution in 1992 (a few years before UAP itself disappeared).

The recognition of handwriting has been one of the projects initiated as early as 1985 by Yvette Chassagne together with the Scientific Council. In cooperation with the French Ministry of Industry when Madame Cresson was Minister, UAP and a 'startup' founded by Physics Nobel Prize laureate Leon Cooper developed a prototype which was presented in 1987 at the Journées Scientifiques of UAP. It was based on neural networks, a subject on which Leon Cooper worked for many years, and operational on PCs of that time, in a context and with an aim different from the A2iA approach. Difficulties in finding industrial partners and a change of policy from the new UAP leadership did not permit further development of the prototype and its production for the professional use in insurance marketing for which it had been designed.

We are happy that Jean-Claude Simon could see the success of his approach in a project which continued work started at the Scientific Council. Together with his family and colleagues we mourn the too early disappearance of a lively and original figure in the scientific community.