Text Segmentation by Automatically Designed Morphological Operators

NINA S. T. HIRATA, JUNIOR BARRERA AND ROUTO TERADA

Instituto de Matemática e Estatística - USP
Rua do Matão, 1010
05508-900 São Paulo - SP - Brazil
<nina,jb,rt>@ime.usp.br

Abstract. Identification of areas corresponding to text in document images is an important step for a character recognition system. In this paper, we briefly review a technique for automatic design of binary morphological operators and show its application to the segmentation of text areas from document page images. We also present an heuristic filter used to refine the segmentation results. Results obtained for two different sets of images are shown.

1 Introduction

Typically, a document image contains objects such as text and figures (diagrams, graphics, half-tones). Information about the localization of these objects in a page image may help document processing systems such as optical character recognition (OCR) systems. The task of separating text from figures is usually composed of two processes : page segmentation and page classification [1]. *Page segmentation* refers to the partitioning of a page image into subregions corresponding to those objects. *Page classification* is the identification of the object type in each region. Several techniques for page segmentation and classification have been proposed. More recent works include [1, 2, 3].

In this paper, we present a novel approach for text segmentation from binary images based on automatically designed binary morphological operators. Automatic design of morphological image operators [4, 5] has been previously applied with success for character recognition [6, 7] and also to other image processing problems. In our approach, a morphological operator is designed for segmenting texts, and its results are further processed by heuristically designed robust morphological filters. The automatically generated segmentation operator plays an important role in the whole segmentation process because its results, even not being perfect, allow a simple post-processing that effectively segments the areas of interest.

Following this, we first review in Section 2 some basic definitions of mathematical morphology for binary images. In Section 3 we recall the automatic design procedure of morphological operators. In Section 4, we present results of the proposed method for several page images, and finally in Section 5 we give the conclusion.

2 Elements of Mathematical Morphology

Let $E = Z^2$ and let $\mathcal{P}(\mathrm{E})$ denote the power set of E. Consider also a finite subset $W \subset E$, containing the origin o of E. The translation of a set $S \subseteq E$ by a vector $z \in E$ is denoted S_z and defined by $S_z = \{y \in E : y = x + z, x \in S\}$. The cardinality of a finite set $S \in \mathcal{P}(\mathrm{E})$ is denoted $|S|$. A mapping $\Psi : \mathcal{P}(\mathrm{E}) \to \mathcal{P}(\mathrm{E})$ is a W-operator if and only if (iff) it is translation-invariant (i.e., $\Psi(S_z) = [\Psi(S)]_z$, for all $z \in E$ and $S \in \mathcal{P}(\mathrm{E})$) and locally defined within W (i.e., $x \in \Psi(S) \Leftrightarrow x \in \Psi(S \cap W_x)$, for all $x \in E$ and $S \in \mathcal{P}(\mathrm{E})$).

Proposition 2.1 *A mapping $\Psi : \mathcal{P}(\mathrm{E}) \to \mathcal{P}(\mathrm{E})$ is a W-operator iff there exists a mapping $\psi : \mathcal{P}(\mathrm{W}) \to \{0, 1\}$ such that*

$$x \in \Psi(S) \Longleftrightarrow \psi(S_{-x} \cap W) = 1$$

for all $x \in E$ and $S \in \mathcal{P}(\mathrm{E})$. The mapping ψ is called the characteristic function of Ψ.

The kernel of a W-operator $\Psi : \mathcal{P}(\mathrm{E}) \to \mathcal{P}(\mathrm{E})$, characterized by a function ψ, is the collection $\mathcal{K}(\Psi) = \{X \in \mathcal{P}(\mathrm{W}) : \psi(X) = 1\}$. In order to design a W-operator, one only needs to design its characteristic function, or specify its kernel.

W-operators can be used for binary image processing because binary images can be regarded as subsets of the image domain (in our case, the set E). Some operators of mathematical morphology for binary images are presented in the remaining of this section.

Definition 2.1 *Let $B \in \mathcal{P}(\mathrm{W})$. The W-operators δ_B and ε_B defined by*

$$x \in \delta_B(S) \Longleftrightarrow S \cap B_x \neq \emptyset$$

0-7695-0878-2/00 $10.00 © 2000 IEEE

and

$$x \in \varepsilon_B(S) \iff B_x \subseteq S$$

for any $x \in E$ *and* $S \in \mathcal{P}(\mathrm{E})$, *are called, respectively,* dilation *and* erosion *by* B. *The set* B *is called a structuring element.*

Definition 2.2 *The operators* γ_B *and* φ_B *from* $\mathcal{P}(\mathrm{E})$ *to* $\mathcal{P}(\mathrm{E})$, *given by*

$$\gamma_B = \delta_B \varepsilon_B$$

and

$$\varphi_B = \varepsilon_B \delta_B,$$

are called, respectively, opening *and* closing *by* B.

Intuitively, one might think of an opening as the operator that removes objects or part of the objects that are smaller than the structuring element, while closing is the operator that fills spaces between the objects that are smaller than the structuring element.

Definition 2.3 *Given* $k > 0$, *the* W*-operator defined by the characteristic function* $\psi(X) = 1 \iff |X| \geq k$, *for all* $X \in \mathcal{P}(\mathrm{W})$, *is the* k-order filter. *If* $k = \lceil |W|/2 \rceil$ *then it is a* median filter.

Let $S \in \mathcal{P}(\mathrm{E})$ and let S^c denote the complement of S. A *hole* of an image S is any finite connected component of S^c (in practice, any connected component of S^c that does not touch the border of the image). The *closing of holes* is an operator that fills all holes of an image. Let $a > 0$. An image operator that removes all components of an image with size smaller than a is called the *a-area open* filter. This filter is useful for removing small objects from an image. Formal definitions of these operators can be found, for instance, in [8, 9].

A W-operator Ψ is anti-extensive iff $\Psi(S) \subseteq S$, for all $S \in \mathcal{P}(\mathrm{E})$. For further details on mathematical morphology the reader may refer to [10, 11, 12].

3 Automatic Design of Morphological Operators

In general, the approaches to design morphological operators consider descriptions of the desired image transformation given in high-level of abstraction. These descriptions are translated into a specification of a morphological operator that realizes the described transformation. The approach in this paper [4, 5] considers pairs of observed-ideal images, like those shown in Fig. 1, as a description of the desired transformation.

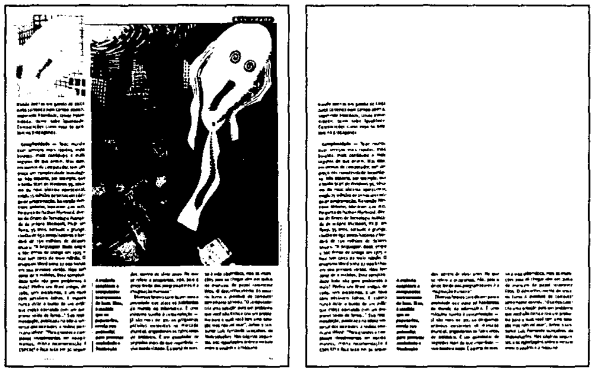


Figure 1: A training pair of observed-ideal images.

The pairs of observed-ideal images (S, I) are considered as realizations of random processes $\mathbf{S}$ and $\mathbf{I}$, and the operators are regarded as estimators of $\mathbf{I}$ in terms of $\mathbf{S}$ [13]. In other words, one would like to find an operator Ψ such that $\Psi(\mathbf{S})$ is as close as possible to $\mathbf{I}$, according to some error measure. Usually, the error measure to be minimized is the mean absolute error, i.e., $MAE\langle\Psi\rangle = E[|\Psi(\mathbf{S})(z) - \mathbf{I}(z)|]$. Under the assumption that the processes $\mathbf{S}$ and $\mathbf{I}$ are jointly stationary, point z is arbitrary. Hence, they can be characterized by a joint process $(\mathbf{X}, \mathbf{y})$ where realizations of $\mathbf{X}$ are subsets of W and realizations of $\mathbf{y}$ are in $\{0, 1\}$. The measure to be minimized is then written as $MAE\langle\Psi\rangle = E[|\psi(\mathbf{X}) - \mathbf{y}|]$. Given the probabilities $P(X) = P(\mathbf{X} = X)$ of observing X and the conditional probabilities $p(1|X) = P(\mathbf{y} = 1|X)$ of observing value 1 in the ideal image given that pattern X has been observed in the observed image, the MAE-optimal W-operator is characterized by the function:

$$\psi_{opt}(X) = \begin{cases} 1, & \text{if } p(1|X) > 0.5, \\ 0, & \text{if } p(1|X) \leq 0.5. \end{cases} \tag{1}$$

In practice, in order to estimate the optimal operator, sample pairs of observed-ideal images are used to estimate the conditional probabilities $p(1|X)$. Some patterns $X \in \mathcal{P}(\mathrm{W})$ may not be observed in the sample and therefore the value of the operator stays undefined for those patterns. Moreover, the representation of the operator by means of its kernel is not adequate in computational terms. Therefore, a learning algorithm [14] is applied to the set of observed patterns. It has two main objectives : (1) to generalize the operator definition for patterns not observed in the sample, and (2) to reduce the representation cost of the operator. In this paper we use the ISI algorithm [4] that meets both objectives. ISI receives as input a collection of examples (X, y) and outputs a collection of intervals that completely describes a W-operator ψ and which is consistent with the input data, i.e., $\psi(X) = y$ for each input pair (X, y).

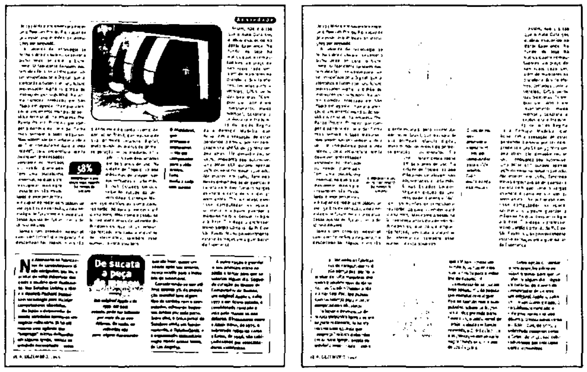


Figure 2: A test page and the result of the designed operator.

Once an operator Ψ_1 is designed from pairs (S, I) of sample images, a second operator Ψ_2 may be designed from pairs $(\Psi_1(S), I)$, a third one from pairs $(\Psi_2(\Psi_1(S)), I)$ and so on. The composed operator $\Psi_2\Psi_1$ is called a two-iteration operator. Analogously, $\Psi_3\Psi_2\Psi_1$ is called a three-iteration operator. Effectiveness of iterated operators has been investigated for the case of increasing operators in [15], and for the class of (not necessarily increasing) W-operators in [16].

4 Experimental Results

The automatic design technique described above was applied to segment text areas from pages of two sets of page images. Training images like those shown in Fig. 1 were used to design the operators. The images were obtained in high resolution (200 and 300 dpi, respectively), binarized by a simple threshold operation, and down-sampled to get images with resolution equivalent to 100dpi. The operators were trained to segment only black characters (in a white background), and only those with standard font size. Since the expected result from the segmentation (ideal image) is a subset of the observed image, we designed anti-extensive operators (all elements of the kernel contains the origin o).

For the first set of images, a two-iteration operator $\Psi = \Psi_2\Psi_1$ were designed using a total of 5 pairs of training images. Letting $(S_1, I_1), \ldots, (S_5, I_5)$ denote the training pairs of images, the first iteration operator, Ψ_1, were designed from $(S_1, I_1), \ldots, (S_4, I_4)$ over the 7×5 window, while the second one, Ψ_2, were designed from $(\Psi_1(S_2), I_2), \ldots, (\Psi_1(S_5), I_5)$ over a 21-point window (the 5×5 window without the four corner points). Figure 2 shows the result of Ψ applied on a test page.

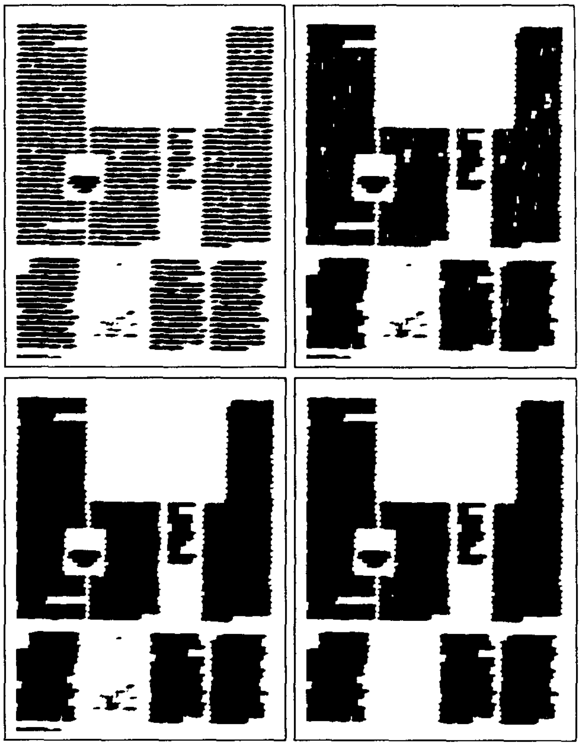


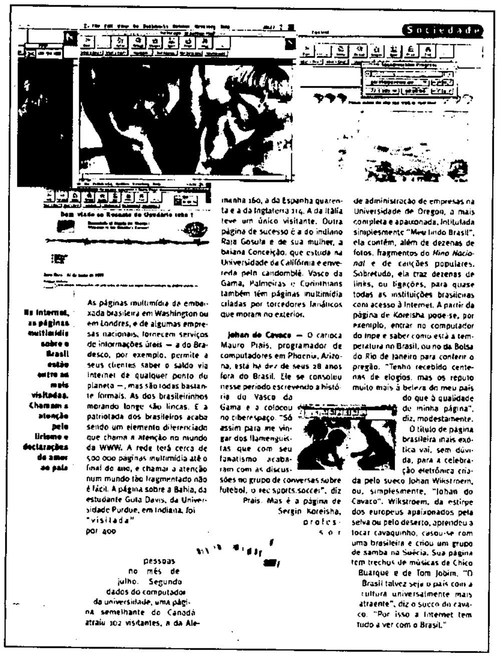
Figure 3: Effects of the post-processing procedure. From left to right, from top to bottom: results of the order filter, of the vertical closing, of the closing of holes, and of the area open, respectively.

Text objects of standard font size are kept almost as they were in the image before the processing, while non-text objects and text of non-standard font size are almost completely removed. Note that the density of pixels in the area corresponding to text objects is much larger than in non-text areas. Based on this fact, a post-processing procedure consisting of the following sequence of filters is applied :

- a 18-order filter, relative to the 7 × 20 window,
- a closing by a vertical line structuring element of length 10,
- the closing of holes operator,
- an area open filter.

The parameters of the order filter and of the closing were empirically adjusted for the whole set of images, while the parameters of the area open filter were adjusted for each page. Figure 3 shows the effect of the post-processing sequence of filters applied on the result shown in Fig. 2.

Next we show several test images (of the first group) and respective segmentation results obtained applying the designed operator plus the post-processing procedure. For each pair, the first image is the original (observed image) while the second one is the respective segmentation result, superposed to the original image. Some pairs are shown in a larger scale for better visualization.



Sociedade

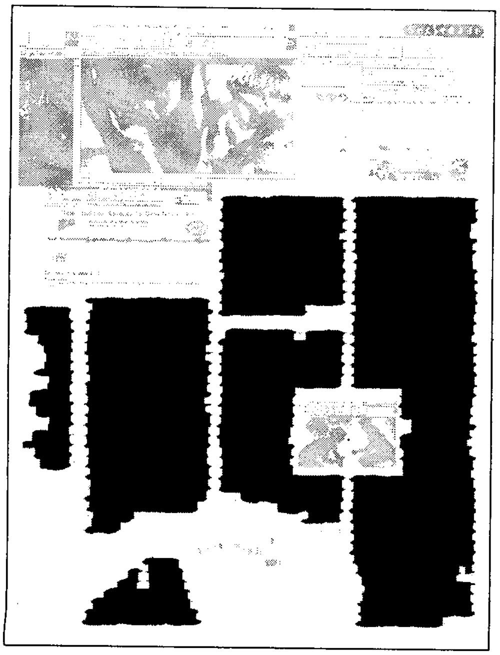
Bem vindo ao Recanto do Operário

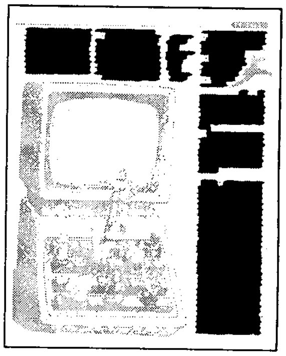
Na Internet, as páginas multimídia sobre o Brasil estão entre as mais visitadas. Chamam a atenção pelo lirismo e declarações de amor ao país

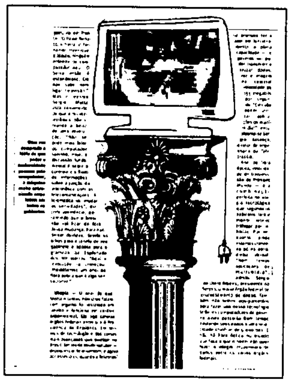
As páginas multimídia da embaixada brasileira em Washington ou em Londres, e de algumas empresas nacionais, fornecem serviços de informações úteis — a do Bradesco, por exemplo, permite a seus clientes saber o saldo via Internet de qualquer ponto do planeta —, mas são todas bastante formais. As dos brasileirinhos morando longe são lindas. E a patriotada dos brasileiros acaba sendo um elemento diferenciado que chama a atenção no mundo da WWW. A rede terá cerca de 500 000 páginas multimídia até o final do ano, e chamar a atenção num mundo tão fragmentado não é fácil. A página sobre a Bahia, da estudante Gula Davis, da Universidade Purdue, em Indiana, foi "visitada" por 400 pessoas no mês de julho. Segundo dados do computador da universidade, uma página semelhante do Canadá atraiu 102 visitantes, a da Alemanha 160, a da Espanha quarenta e a da Inglaterra 314. A da Itália teve um único visitante. Outra página de sucesso é a do indiano Raja Gosula e de sua mulher, a baiana Conceição, que estuda na Universidade da Califórnia e enveredа pelo candomblé. Vasco da Gama, Palmeiras e Corinthians também têm páginas multimídia criadas por torcedores fanáticos que moram no exterior.

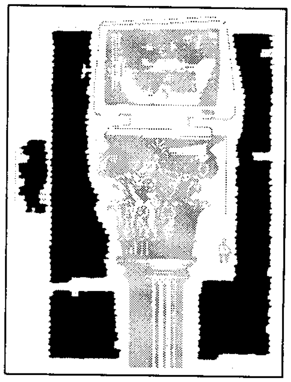
Johan do Cavaco — O carioca Mauro Prais, programador de computadores em Phoenix, Arizona, está há dez de seus 28 anos fora do Brasil. Ele se consolou nesse período escrevendo a história do Vasco da Gama e a colocou no ciberespaço. "Só assim para me vingar dos flamenguistas que com seu fanatismo acabaram com as discussões no grupo de conversas sobre futebol, o rec.sports.soccer", diz Prais. Mas é a página de Sergio Koreisha, professor de administração de empresas na Universidade de Oregon, a mais completa e apaixonada. Intitulada simplesmente "Meu lindo Brasil", ela contém, além de dezenas de fotos, fragmentos do *Hino Nacional* e de canções populares. Sobretudo, ela traz dezenas de links, ou ligações, para quase todas as instituições brasileiras com acesso à Internet. A partir da página de Koreisha pode-se, por exemplo, entrar no computador do Inpe e saber como está a temperatura no Brasil, ou no da Bolsa do Rio de Janeiro para conferir o pregão. "Tenho recebido centenas de elogios, mas os reputo muito mais à beleza do meu país do que à qualidade de minha página", diz, modestamente.

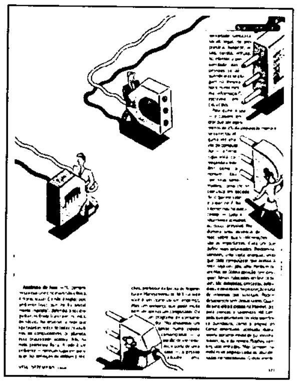
O título de página brasileira mais exótica vai, sem dúvida, para a celebração eletrônica criada pelo sueco Johan Wikstroem, ou, simplesmente, "Johan do Cavaco". Wikstroem, da estirpe dos europeus apaixonados pela selva ou pelo deserto, aprendeu a tocar cavaquinho, casou-se com uma brasileira e criou um grupo de samba na Suécia. Sua página tem trechos de músicas de Chico Buarque e de Tom Jobim. "O Brasil talvez seja o país com a cultura universalmente mais atraente", diz o sueco do cavaco. "Por isso a Internet tem tudo a ver com o Brasil."

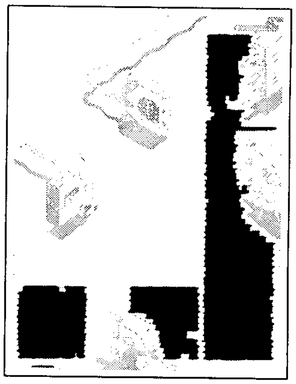


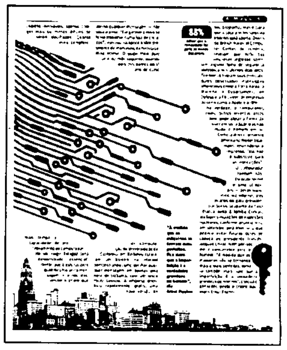


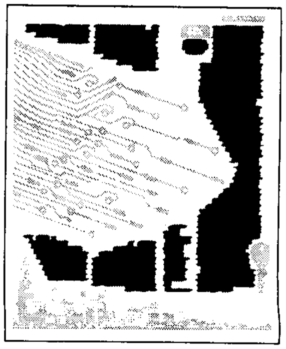


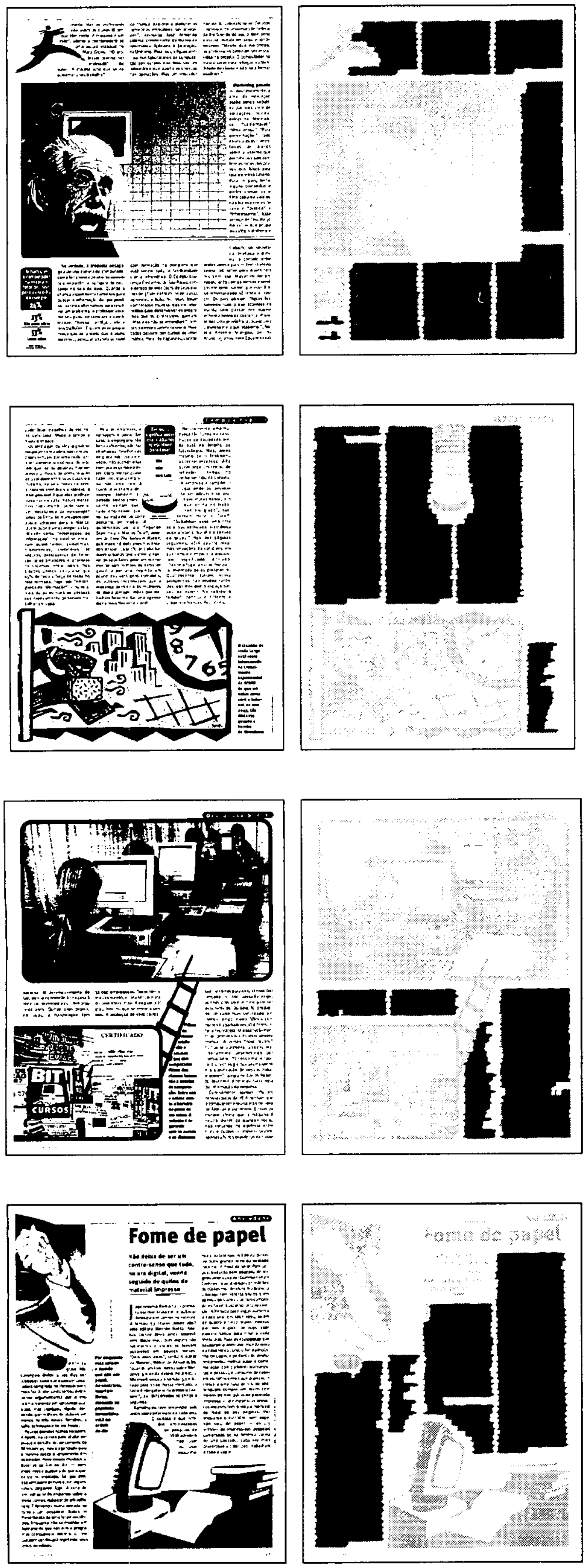












O usam computador, contra quase 35% da classe A/B. Intuitivamente, as classes menos favorecidas sabem que, no fundo, não há máquina capaz de mudar as desigualdades sociais. Isso é tarefa do homem. "O computador explicita as desigualdades, mas não as cria", sustenta o americano Robert J. Samuelson. "Depende do que você faz com ele. É um pouco como o fogão caseiro, que todo mundo tem — alguns fritam hambúrgueres, outros fazem manjares de deuses." Só que o Brasil ainda está na era do pré-hambúrguer: apenas um em cada 100 brasileiros tem acesso a um computador. Enquanto nos Estados Unidos já se contabilizam indicadores profissionais gerados pela informatização — o tecnotrabalhador recebe em média 10% a 15% mais do que o trabalhador americano convencional —, no Brasil o mundo das oportunidades ainda não se deslocou para baixo. "Trabalhador informado não fica desempregado", prometia o Ministério da Educação nos anos 70, ao lançar a campanha de alfabetização para adultos, o Mobral. De lá para cá, avançou-se pouco para arrancar o país da colocação de lanterninha latino-americano em matéria de alfabetização. Ainda hoje, um em cada cinco brasileiros com mais de 15 anos é analfabeto, contra um em 25 na Argentina e um em nove no México. Hoje, a frase do Mobral deveria começar com "Um homem e uma mulher informatizados...", dado que o mercado de trabalho abriu suas comportas para a mão-de-obra feminina, e alfabetização para efeitos de arrumar emprego, há muito deixou de significar saber o alfabeto e as quatro operações.

A busca de uma instrução acima do 1º grau, condição para entrar em empresas como a Volkswagen, levou milhares de brasileiros para o exame de madureza, o vestibular

Ter computador significa ter importância social:

53%

63%
Classe B

57%
Classe C

37%
Classe A/B

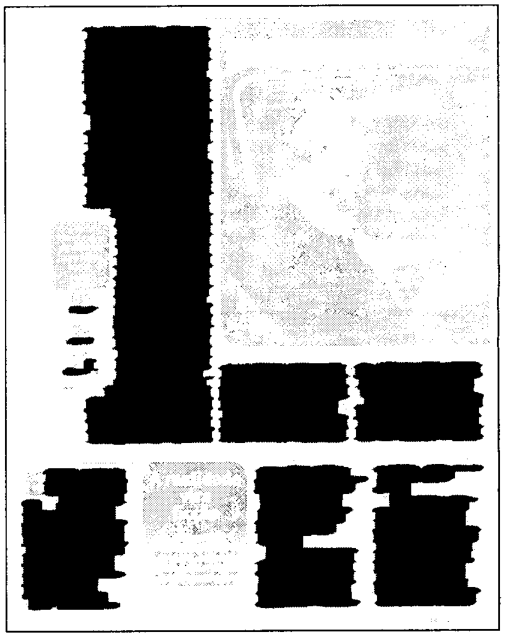
A realidade vira ficção

Entre os 110 000 alunos da Data Control em 1995, um era virtual — "Lucas", o adolescente-problema vivido pelo ator Pedro Vasconcelos em *A Próxima Vítima*, a novela das 8 encerrada poucas semanas atrás. O enquadramento social imaginário do aprendizado da informática foi eficaz. Junto com "Lucas", que graças ao curso voltou a trilhar o bom caminho, freqüentava as aulas sua namorada, "Iara", que de início falava "pobrema", mas acaba arrumando um bom emprego num frigorífico.

Era a primeira vez que uma empresa de cursos de computação comprava espaço nobre de merchandising em novela, disputando a trama com sandálias Azaléia, cerveja Antarctica e carros da Ford.

"A informática, como a novela, já faz parte da rotina dos brasileiros", diz Giulia Caleffi, gerente de mídia da Data Control, que gasta 1,5 milhão de reais por ano em marketing. O merchandising, que durou dois meses, coincidiu com a inauguração da primeira filial da

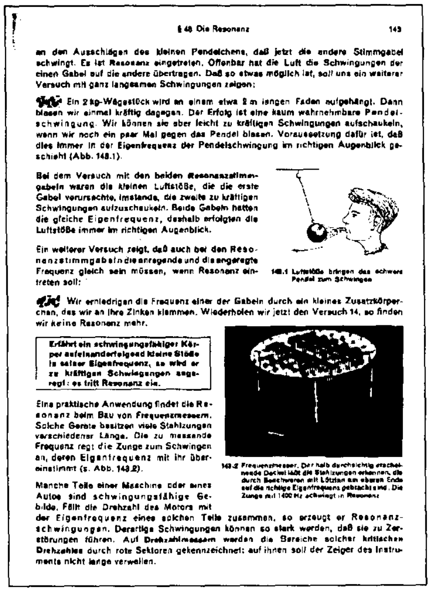
VEJA DEZEMBRO 1995



In all these pages, only few errors were verified. In one of the pages, the page number was missed by the post-processing procedure, because the object was smaller than other non-desired components removed by the area open filter. Some text in italic font were not segmented because they were not present in the training images.

The second set of images was obtained from a different source. The images in this set have, in general, a more complex layout than the ones in the first set.

For instance, they include mathematical formulas and their figures contain text of standard size. For this set, a three-iteration operator has been designed using a total of 8 pairs of training images, over the 7×5, 5×5 and 21-point (the 5×5 window without the four corner points) windows, respectively. For the first-iteration operator, 6 pairs of training images were used, while for the last two iterations all 8 pairs were used. The same post-processing procedure described before, with the same parameter values, was applied to the resulting images. Next we show the segmentation result for some test pages of the second set.



§ 48 Die Resonanz 143

an den Ausschlägen des kleinen Pendelchens, daß jetzt die andere Stimmgabel schwingt. Es ist Resonanz eingetreten. Offenbar hat die Luft die Schwingungen der einen Gabel auf die andere übertragen. Daß so etwas möglich ist, soll uns ein weiterer Versuch mit ganz langsamen Schwingungen zeigen:

Ein 2 kg-Wägestück wird an einem etwa 2 m langen Faden aufgehängt. Dann blasen wir einmal kräftig dagegen. Der Erfolg ist eine kaum wahrnehmbare Pendelschwingung. Wir können sie aber leicht zu kräftigen Schwingungen aufschaukeln, wenn wir noch ein paar Mal gegen das Pendel blasen. Voraussetzung dafür ist, daß dies immer in der Eigenfrequenz der Pendelschwingung im richtigen Augenblick geschieht (Abb. 148.1).

Bei dem Versuch mit den beiden Resonanzstimmgabeln waren die kleinen Luftstöße, die die erste Gabel verursachte, imstande, die zweite zu kräftigen Schwingungen aufzuschaukeln. Beide Gabeln hatten die gleiche Eigenfrequenz, deshalb erfolgten die Luftstöße immer im richtigen Augenblick.

Ein weiterer Versuch zeigt, daß auch bei den Resonanzstimmgabeln die anregende und die angeregte Frequenz gleich sein müssen, wenn Resonanz eintreten soll:

148.1 Luftstöße bringen das schwere Pendel zum Schwingen

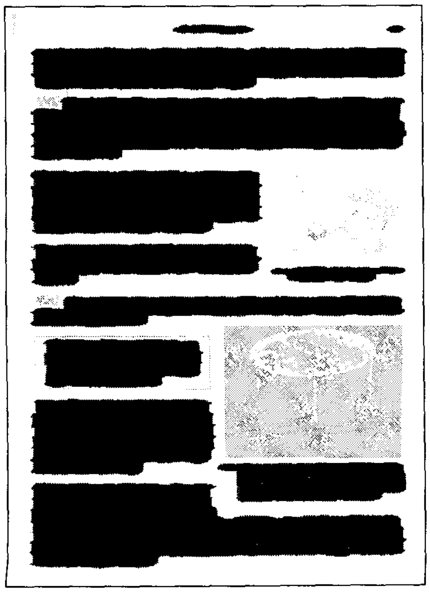
Wir erniedrigen die Frequenz einer der Gabeln durch ein kleines Zusatzkörperchen, das wir an ihre Zinken klemmen. Wiederholen wir jetzt den Versuch 14, so finden wir keine Resonanz mehr.

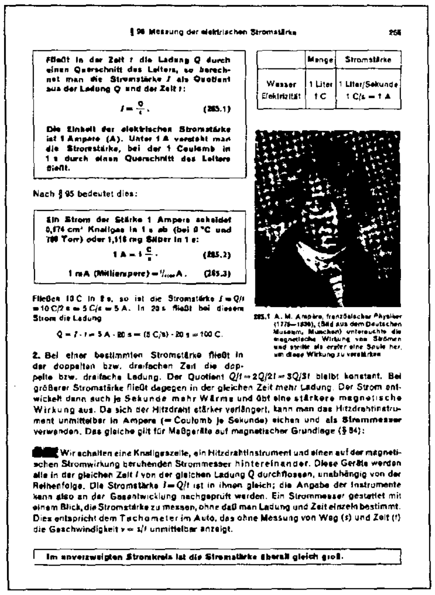
Erfährt ein schwingungsfähiger Körper aufeinanderfolgend kleine Stöße in seiner Eigenfrequenz, so wird er zu kräftigen Schwingungen angeregt: es tritt Resonanz ein.

Eine praktische Anwendung findet die Resonanz beim Bau von Frequenzmessern. Solche Geräte besitzen viele Stahlzungen verschiedener Länge. Die zu messende Frequenz regt die Zunge zum Schwingen an, deren Eigenfrequenz mit ihr übereinstimmt (s. Abb. 143.2).

143.2 Frequenzmesser. Der halb durchsichtig erscheinende Deckel läßt die Stahlzungen erkennen, die durch Beschweren mit Lötzinn am oberen Ende auf die richtige Eigenfrequenz gebracht sind. Die Zunge mit 1400 Hz schwingt in Resonanz

Manche Teile einer Maschine oder eines Autos sind schwingungsfähige Gebilde. Fällt die Drehzahl des Motors mit der Eigenfrequenz eines solchen Teils zusammen, so erzeugt er Resonanzschwingungen. Derartige Schwingungen können so stark werden, daß sie zu Zerstörungen führen. Auf Drehzahlmessern werden die Bereiche solcher kritischen Drehzahlen durch rote Sektoren gekennzeichnet; auf ihnen soll der Zeiger des Instruments nicht lange verweilen.





§ 96 Messung der elektrischen Stromstärke 255

Fließt in der Zeit t die Ladung Q durch einen Querschnitt des Leiters, so berechnet man die Stromstärke I als Quotient aus der Ladung Q und der Zeit t:

$$I = \frac{Q}{t}. \qquad (255.1)$$

Die Einheit der elektrischen Stromstärke ist 1 Ampere (A). Unter 1 A versteht man die Stromstärke, bei der 1 Coulomb in 1 s durch einen Querschnitt des Leiters fließt.

	Menge	Stromstärke
Wasser Elektrizität	1 Liter 1 C	1 Liter/Sekunde 1 C/s = 1 A

Nach § 95 bedeutet dies:

Ein Strom der Stärke 1 Ampere scheidet 0,174 cm³ Knallgas in 1 s ab (bei 0 °C und 760 Torr) oder 1,118 mg Silber in 1 s:

$$1\,\mathrm{A} = 1\,\frac{\mathrm{C}}{\mathrm{s}}. \qquad (255.2)$$

$$1\,\mathrm{mA}\ (\text{Milliampere}) = {}^{1}/_{1000}\,\mathrm{A}. \qquad (255.3)$$

Fließen 10 C in 2 s, so ist die Stromstärke $I = Q/t = 10\,\mathrm{C}/2\,\mathrm{s} = 5\,\mathrm{C/s} = 5\,\mathrm{A}$. In 20 s fließt bei diesem Strom die Ladung

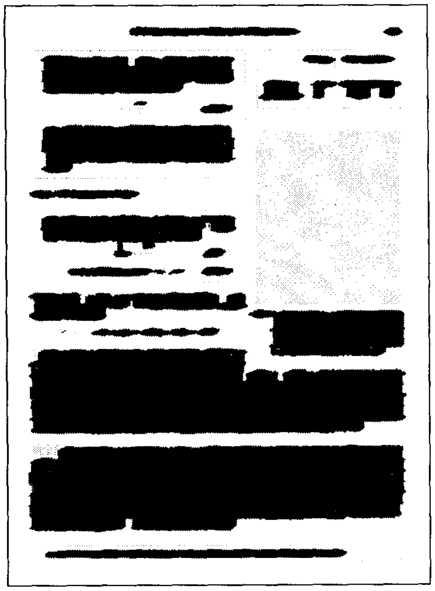
$$Q = I \cdot t = 5\,\mathrm{A} \cdot 20\,\mathrm{s} = (5\,\mathrm{C/s}) \cdot 20\,\mathrm{s} = 100\,\mathrm{C}.$$

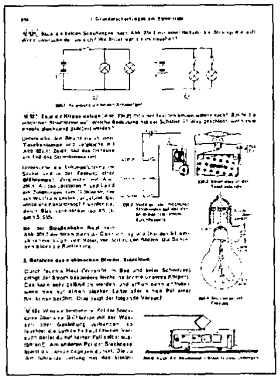
255.1 A. M. Ampère, französischer Physiker (1775—1836), (Bild aus dem Deutschen Museum, München) untersuchte die magnetische Wirkung von Strömen und stellte als erster eine Spule her, um diese Wirkung zu verstärken

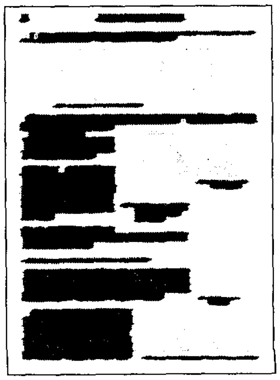
2. Bei einer bestimmten Stromstärke fließt in der doppelten bzw. dreifachen Zeit die doppelte bzw. dreifache Ladung. Der Quotient $Q/t = 2Q/2t = 3Q/3t$ bleibt konstant. Bei größerer Stromstärke fließt dagegen in der gleichen Zeit mehr Ladung. Der Strom entwickelt dann auch je Sekunde mehr Wärme und übt eine stärkere magnetische Wirkung aus. Da sich der Hitzdraht stärker verlängert, kann man das Hitzdrahtinstrument unmittelbar in Ampere (= Coulomb je Sekunde) eichen und als Strommesser verwenden. Das gleiche gilt für Maßgeräte auf magnetischer Grundlage (§ 84):

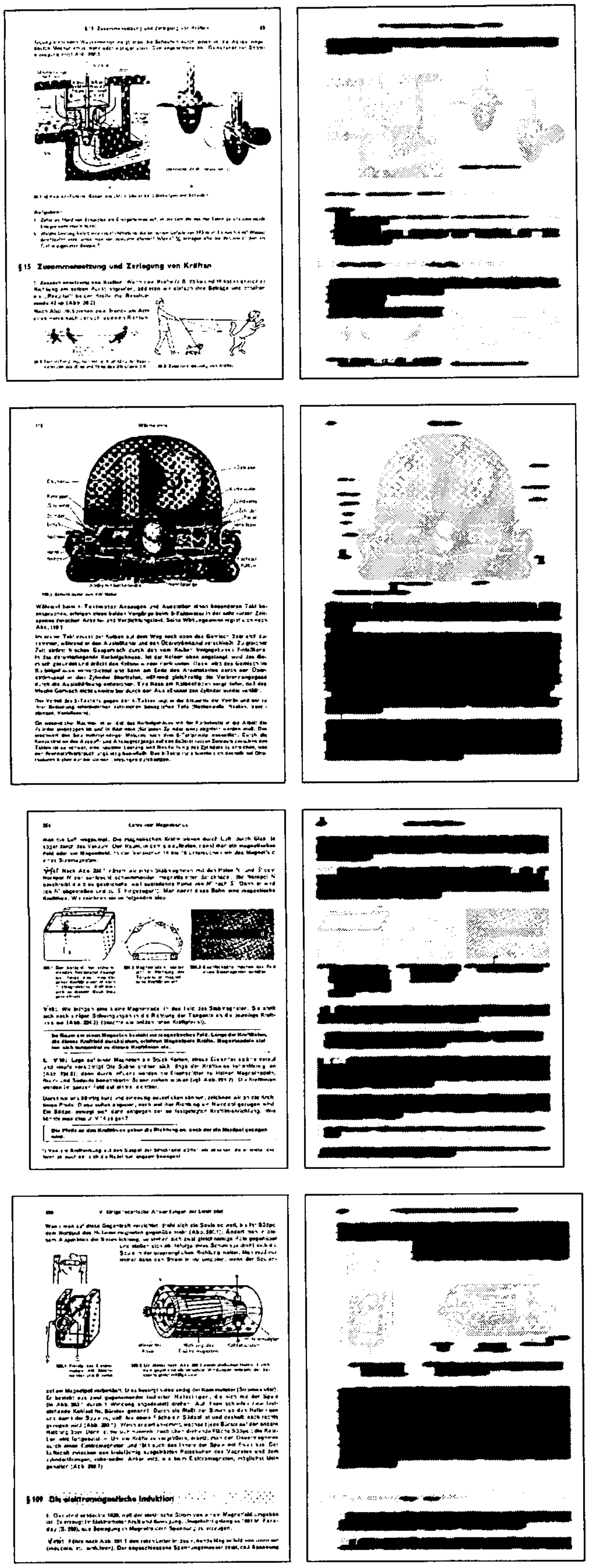
Wir schalten eine Knallgaszelle, ein Hitzdrahtinstrument und einen auf der magnetischen Stromwirkung beruhenden Strommesser hintereinander. Diese Geräte werden alle in der gleichen Zeit t von der gleichen Ladung Q durchflossen, unabhängig von der Reihenfolge. Die Stromstärke $I = Q/t$ ist in ihnen gleich; die Angabe der Instrumente kann also an der Gasentwicklung nachgeprüft werden. Ein Strommesser gestattet mit einem Blick, die Stromstärke zu messen, ohne daß man Ladung und Zeit einzeln bestimmt. Dies entspricht dem Tachometer im Auto, das ohne Messung von Weg (s) und Zeit (t) die Geschwindigkeit $v = s/t$ unmittelbar anzeigt.

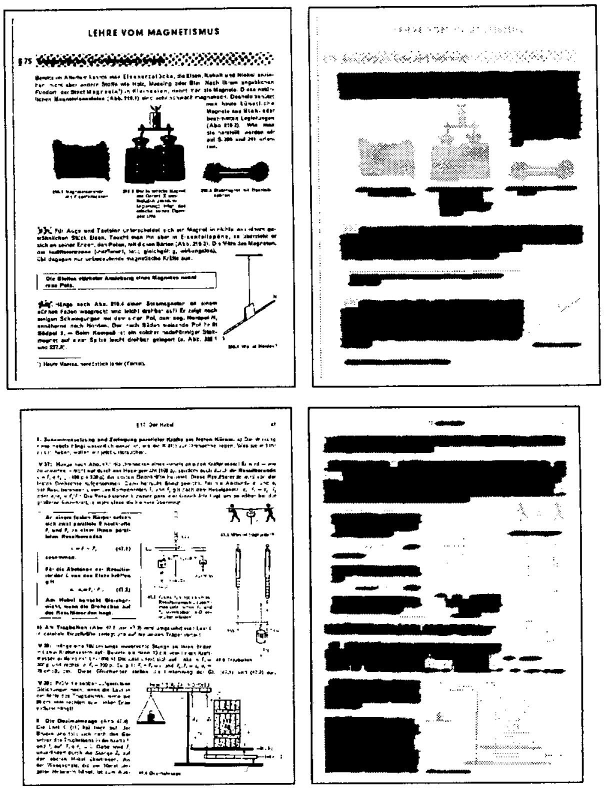
Im unverzweigten Stromkreis ist die Stromstärke überall gleich groß.











In the second set of images, the results are good, but some page numbers and formulas were missed. Some text inside the diagrams were segmented while others were not. This may be due to the fact that they were regarded as non-text elements for the design of the operator. Better results should be obtained if more training data is used in conjunction with careful selection of the objects to be segmented.

5 Conclusion

Very good segmentation results have been obtained by the method proposed in this paper. Some advantages of our method are: (1) it does not require any hypothesis about the characteristics of the objects present in the page images; only a relatively small number of observed-ideal pair of images is required for training an operator, (2) we have observed that the designed operator is robust with relation to slightly skewed images.

The heuristic post-processing procedure we used seems to be robust, only with the inconvenience that the parameter of the area open filter must be adjusted for each image. Some filters that produce markers for the objects of interest (so to avoid the adjustment of the area open filter) have been tested (not shown). It is noteworthy to mention that it may be possible to replace the post-processing procedure by automatic designed operators, eliminating at all the need for adjusting the parameters of the post processing procedure. This is one of the issues for future research.

Another issue we would like to investigate further is the appropriate resolution of the images. When documents are to be processed by OCRs they are usually scanned at a relatively high resolution (300dpi or more). Since time to process an image is also proportional to the image size, a way to cut down the processing time is to use low resolution images. In our approach, we have used images of resolution approximately 100dpi, and one of our next goals is to investigate the application of the method on images of lower resolutions.

6 Acknowledgements

N. S. T. Hirata acknowledges support from FAPESP (process 98/14328-6).

References

[1] T. Pavlidis. Page Segmentation and Classification. *CVGIP: Graphical Models and Image Processing*, 54(6):484-496, November 1992.

[2] K. Kise, A. Sato, and M. Iwata. Segmentation of Page Images Using the Area Voronoi Diagram. *Computer Vision and Image Understanding*, 70(3):370 382, June 1998.

[3] A. Antonacopoulos. Page Segmentation Using the Description of the Background. *Computer Vision and Image Understanding*, 70(3):350-369, June 1998.

[4] J. Barrera, E. R. Dougherty, and N. S. Tomita. Automatic Programming of Binary Morphological Machines by Design of Statistically Optimal Operators in the Context of Computational Learning Theory. *Electronic Imaging*, 6(1):54-67, January 1997.

[5] J. Barrera, R. Terada, R. Hirata Jr, and N. S. T. Hirata. Automatic Programming of Morphological Machines by PAC Learning. *Fundamenta Informaticae*, 41(1-2):229-258, January 2000.

[6] J. Barrera, R. Terada, F. S. C. da Silva, and N. S. Tomita. Automatic Programming of Morphological Machines for OCR. In *Mathematical Morphology and its Applications to Image and Signal Processing*, pages 385-392, Atlanta, GA, May 1996. International Symposium on Mathematical Morphology, Kluwer Academic Publishers.

[7] J. Barrera, R. Terada, R. A. Lotufo, N. S. T. Hirata, R. Hirata Jr., and F. A. Zampirolli. An OCR based on Mathematical Morphology. In *Nonlinear Image Processing IX*, volume 3304 of *Proceedings of SPIE*, pages 197-208, San Jose, CA, January 1998.

[8] J. Barrera, G. J. F. Banon, R. A. Lotufo, and R. Hirata Jr. MMach: a Mathematical Morphology Toolbox for the Khoros System. *Electronic Imaging*, 7(1):174-210, 1998.

[9] P. Salembier and J. Serra. Flat Zones Filtering, Connected Operators, and Filters by Reconstruction. *IEEE Transactions on Image Processing*, 4(8):1153-1160, August 1995.

[10] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.

[11] H. J. A. M. Heijmans. *Morphological Image Operators*. Academic Press, Boston, 1994.

[12] P. Soille. *Morphological Image Analysis*. Springer-Verlag, Berlin, 1999.

[13] E. R. Dougherty. *Random Processes for Image and Signal Processing*. SPIE and IEEE Presses, Bellingham, 1998.

[14] T. M. Mitchell. *Machine Learning*. McGraw-Hill Series in Computer Science. McGraw-Hill, March 1997.

[15] E. R. Dougherty, Y. Zhang, and Y. Chen. Optimal Iterative Increasing Binary Morphological Filters. *Optical Engineering*, 35(12):3495-3507, December 1996.

[16] N. S. T. Hirata, E. R. Dougherty, and J. Barrera. Iterative Design of Morphological Binary Image Operators. *to appear*, 2000.