Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

Document Analysis Systems that Improve with Use

George Nagy

Received: ?

Abstract Document analysis tasks for which representative labeled training samples are available have been largely solved. The next frontier is coping with hitherto unseen formats, unusual typefaces, idiosyncratic handwriting and imperfect image acquisition. Adaptive and style-constrained classification methods can overcome some expected variability, but human intervention will remain necessary in many tasks. Interactive pattern recognition includes data exploration and active learning as well as access to stored documents. The principle of "green interaction" is to make use of every intervention to reduce the likelihood that the automated system will make the same mistake again and again. Some of these techniques may pop up in forthcoming personal camera-based memex-like applications that will have a far broader range of input documents and scene text than the current, successful but highly specialized, systems for patents, postal addresses, bank checks and books.

Keywords interactive document analysis adaptive classification style constrained recognition camera-based OCR memex lifetime reader

1. Introduction

This is a review of a lifetime of industrial and academic research on character and document recognition. The quest is far from over. Perhaps some reflections on my own successes and failures can provide inspiration for research on approaches beyond those underlying the many successful applications where the potential for further improvement has already plateaued. For the next step, I propose a reincarnation of Vannevar Bush's *Memex* [1], the Lifetime Reader (§5).

Because I am reviewing over fifty years of work [2], I ask your indulgence for any sense of déjà vu. I also feel obliged to emphasize that this is not a balanced survey. I shall zig and zag shamelessly in the orthogonal directions of human-machine interaction and autonomous adaptation. For a more aseptic review, please see [3].

G. Nagy
Rensselaer Polytechnic Institute
Troy, NY 12180, USA
e-mail: nagy@ecse.rpi.edu

The traditional problem of pattern recognition and machine learning is to classify a set of objects characterized by feature vectors (single pixels or combinations of pixels) into mutually exclusive classes. In our domain the objects are digitized artifacts created for symbolic and semiotic communications across time and space, i.e., *documents* and *signs*.

In supervised classification, a subset of the objects—ideally a representative random sample, but often only a convenience sample—is manually labeled for use as a training set of for estimating the parameters required for automatic assignment of labels to the remaining samples. The labels may be symbols of some alphabet, logographs, words or phrases, font or writer identities, document components (e.g., *title, author, figure caption, citation, table*), or document categories *(letter, invoice, advertisement, duplicate, forgery*). Fig. 1 is a 1960's style illustration of this paradigm applied to character recognition.

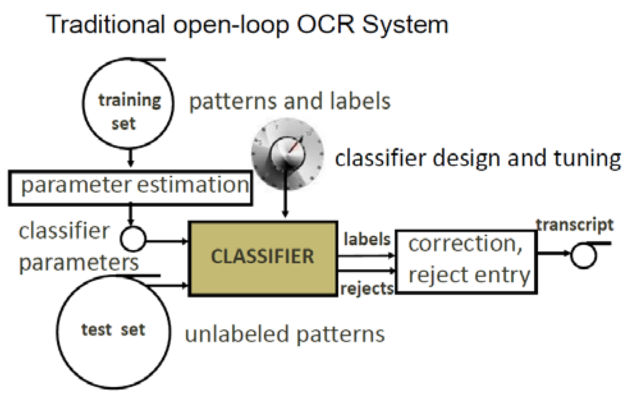


Fig. 1. The classifier is trained on labeled patterns. After classifying the unlabeled test set (operational data), an operator corrects errors and labels rejects to produce a usable transcript.

Data processed operationally is always much larger and necessarily more representative of itself than the training sets used for design. We should therefore draw on the operational data stream for improving the classifier by interactive labeling and training or by autonomous adaptation. Less direct feedback could also be provided by error-sensitive downstream client-systems when they try to make sense of the processed documents.

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

Regardless of whether human-labeled or machine-labeled samples (or both) are used to refine the classifier parameter estimates, the underlying assumption is the consistency of each batch. Manual or automated demarcation of batches of similar samples was unnecessary in conventional OCR post-processing where errors were merely corrected but operator intervention was not used to improve the essentially *static* (in earlier times, hard-wired) OCR engines. It is only recently that large-capacity classifiers like Deep Neural Networks obviated the need to readjust the classifier parameters for any change of document font, layout, or language even if it has already been seen in the past. Their output on many documents is still far from perfect.

Instead of starting out by labeling some samples for initializing the classifier, the samples can be clustered into groups such that the within-group distances between the feature vectors are smaller than the between-group distances. This is called *unsupervised* classification or *l*earning *without a teacher*, but still requires specifying some cluster size, shape, distribution, cardinality constraint or other optimization criterion (*hyper-parameters*). Grouping similar samples minimizes human effort because only one or a few representative members of each cluster need to be manually labeled [4].

Clustering the feature vectors of character images also opens the way for labeling by a computer agent, for instance by assigning an alphabetic label to each cluster according to symbol co-occurrence frequencies, or a dictionary, or some higher-level language model [5,6,7 8]. This works admirably regardless of typeface, as long as the entire text is in the same font and language. It is like solving a substitution cypher. Crypt-OCR is appealing and could be more fully exploited in combination with other methods.

These topics are all interdependent and jumbled in my mind, yet I must try to present them sequentially. In Section 2, I discuss several kinds of interaction that are relevant to DIA, but omit others where I have little or no personal experience, like CAPTCHAs, word-completion, line-following for vectorization, interactive mathematical proofs, and social networks. In Section 3, I offer my two-penny's worth on adaptation, which has a huge literature, and on style constraints, which so far has been of interest mainly to myself, my students, and a few gullible researchers. Illustrations of concrete examples from our experiments are reproduced in Section 4. Section 5 sketches the Lifetime Reader, a prospective application that could make use of some notions discussed in the preceding sections. Perhaps some of you can help the Lifetime Reader reach operational status before my memory leaks get much worse. I have not yet reached any immutable Conclusions, so I will stop there.

2. Interactive Pattern Recognition

Survival of the species demands expertise in pattern recognition. So we can take it for granted that we are all experts who can help machines to recognize or classify objects. We readily acknowledge that computers have better memory and are faster than we are. We therefore accept examining only small subsets of data displayed by the computer and will gladly let the computer do the dog work according to our insightful instructions. Dorothea Blostein and I reported some figures on the cost of operator time relative to overall document processing cost [9].

There are many ways of improving automated data analysis based on selective displays of relevant aspects of the data or of the results of the automated process on previous data. J. Zou and I reviewed some of the motivation and benefits of human-computer interaction (HCI) for pattern recognition in [10]. Here I attempt to cluster my remarks into four categories: Exploratory Data Analysis, Performance Evaluation, Active Sampling, and Green Interaction.

2.1 Exploratory Data Analysis

The toolkits of Exploratory Data Analysis (EDA) help to inspect datasets characterized by multi-dimensional descriptors (features) in order to determine what type of automated processing (dimensionality reduction, feature selection, classification, identity confirmation …) would yield the best results. Typical tools are sample selection, projections into low-dimensional spaces with additional dimensions represented by color or distinctive glyphs, plots of marginal or low-dimensional joint probability distributions, and descriptive statistics like minima, maxima, averages, correlation coefficients, and higher-order expectations of selected samples.

Early work in EDA was reviewed in a book by Y.T. Chien [11]. Influential proposals were set forth in 1970 by Ball and Hall [12] and Sammon [13], followed a few years later by Tukey's still useful book on data analysis [14]. Parts of Gelsema's ISPAHAN were commercialized for interactive blood cell analysis [15,16]. A thoughtful examination of the various steps in EDA (perhaps slightly biased towards a central role for clustering) appeared in Jain's and Dubes' 1988 *Algorithms for Clustering Data* [17], and in the same year Siedlecki, Siedlecka and Sklansky published their survey [18]. By the end of the century, EDA was morphing into Data Visualization (DV) [19]. A set of modern and powerful java-based EDA tools (*Mirage*) were crafted by T.K. Ho and her colleagues at Bell Labs [20].

Although technological advances in computer graphics (frame-rate, resolution, color-depth, stereo, holograms, and virtual reality) promote further progress in exploratory data analysis and data visualization, insight into the configurations of samples of several classes and subclasses in feature spaces of several dozen or several hundred dimensions remains elusive. Many of our insights from two and three dimensions are misleading [21]. This could be expected from extrapolating Flatland [22].

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

2.2 Performance Evaluation

How can we evaluate conventional, interactive and adaptive pattern recognition systems? We are often biased in assessing the recognition results of systems that we have designed, redesigned and tested and re-tested over a period of years—usually with the same data. In 1983 I collected some common instances of statistical malfeasance in a two-page *PAMI* article (my favorite), with references omitted to protect the guilty [23]. Some of my harangues against subtle ways of training on the test set or optimistic assumptions of independence were recently reprised by others [24].

Later on, during three stints at ISRI-UNLV, I learned a great deal about OCR benchmarking from Tom Nartker and Steve Rice. I summarized what I knew about performance evaluation in [25], and the three of us compiled and annotated OCR errors in a coffee-table book that never made the NYT best-sellers list [26]. To my regret, I did not mention the useful McNemar's hypothesis test for comparing classifiers when the identity of the misrecognized objects is known [27].

What is the role of Henry Baird's document defect models [28] in training and testing OCR systems? Dan Lopresti and I serendipitously gave complementary, back-to-back papers on the subject at an ISRI conference. We recruited some younger and brighter helpers and combined our ideas in a PAMI article [29]. We later contributed to defect models the single entirely unavoidable source of noise in scanning, random-phase spatial sampling, that had previously been of interest only in satellite remote sensing [30]. A current alternative for augmenting training sets are Generative Adversarial Neural Networks (GANNs)[31]

In the absence of useful feedback from downstream programs, human correction time (using an appropriate GUI) is a good measure of classifier performance. Unfortunately it turns out that humans are far from infallible or even consensual in labeling document artifacts [32,33]. Lamiroy and Lopresti propose a comprehensive benchmarking system that accommodates multiple ground truths and offers a fresh selection of samples for each experiment for credible performance comparisons [34]. My mind reels at the intricacy of interaction, adaptation, and evaluation of distributed DIA applications, each processing different streams of data for different clients.

2.3 Active Sampling

The key idea of active sampling/learning is to label only samples too close to the current classification boundary for a reliable decision [35,36]. Reducing the size of the training set offers two advantages: less operator time for labeling, and less machine time for training the classifier. However, we may be picketing irrelevant part of the sample space because the exact location of the boundary is unknown until the classifier is fully trained. Proposed information-theoretic selection criteria are subject to the same caveat [37].

Active sampling has also been applied to selecting features, based on their predictive capability, with minimum acquisition costs. The candidate features are extracted from a region of the sampling space where the choice of features is likely to have the greatest impact [38,39].

2.4 Green Interaction

Green interaction in OCR consists either of proofreading the entire output of the classifier, or correcting only rejected (unlabeled) character samples, and then retraining the classifier with the re-labeled data. The classify-label cycle is then iterated on new data. I dubbed this "green interaction" in 2011 [40]. This notion is illustrated in Fig. 2. Green interaction in an industrial OCR engine is described in [59]. Examples from our own work appear in Section 4.

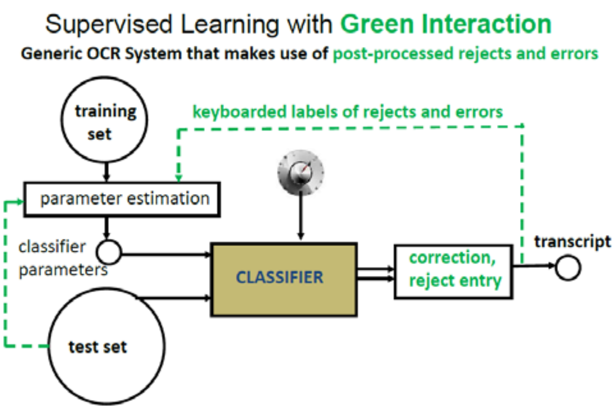


Fig. 2. Green interaction recycles rather than using human corrections only to correct the current output (figures in color on line)

Parsimonious human interaction throughout the transcription process is better than intervention only at the start and end. The question that arises naturally is: After how many samples should one switch from automated classification to manual labeling and vice-versa? We proposed a solution for optimal splits of the data into training and test sets in [41]. Although we provided some supportive experimental evidence, our theory is based on earlier results on learning rates and some questionable assumptions. The efficient use of human labor in operational settings requires more substantial research.

3. Adaptation and Style Constrained Classification

Adaptation means different things to different people. By *adaptive classification* I mean improved classification of a set of unlabeled patterns by taking advantage of other unlabeled patterns. Adaptive and style-constrained classification work best when the entire set of patterns – from the same page, book, or same-source letters - exhibits font, printer, scanner, or vocabulary consistency.

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

Glen Shelton's and my 1965 experiments on adaptive recognition of typewritten characters followed the footsteps of Gold (MAUDE adapted to the lengths of Morse code spaces, dots and dashes [42]), Cooper and Cooper (adaptive signal detection [43]), Fralick [44], Braverman [45], Dorofeyuk [46], Scudder [47], Lucky (adaptive equalization [48,49]) and many others. The difficulties inherent in the various approaches were analyzed by Spragins in his 1966 survey [501 and in a 1971 volume edited by Ya. Z. Tsypkin [51], who was a notable early contributor [52]. These notions were the forerunners of mixture identification models based on Expectation Maximization. Some of the terminology has changed, but *stochastic approximation* (Robbins-Munro algorithm), *self-training / -adaptation / -organization, learning without a teacher / trainer, inductive transfer, co-training, semi-supervised / unsupervised / transfer / surrogate* and *decision-directed learning* remain hot topics.

3.1 An Adaptive Classifier

(a)

(b)

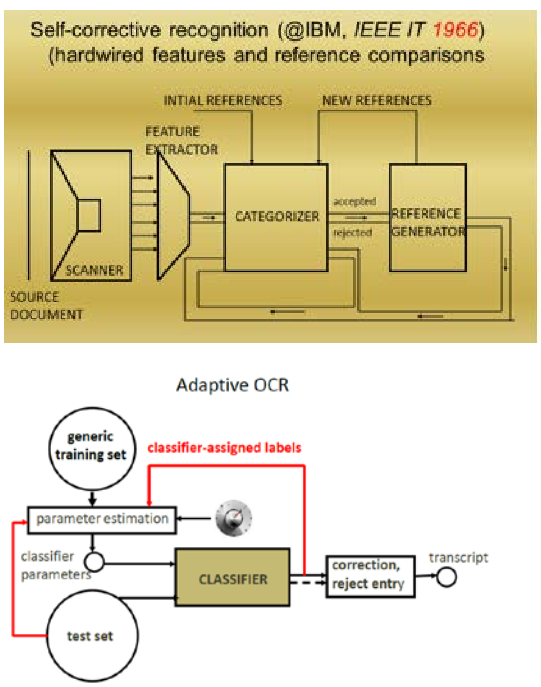


Fig. 3. Adaptive character recognition (a) As implemented at IBM (results in Fig. 7), and (b) as an example of a more general adaptive paradigm.

Our 1966 character recognition system extracted n-tuple features by specialized hardware [53]. The 7-9 pixel n-tuples were specified by plug-boards designed for chain printers. Fig 3a shows the original scheme. It is redrawn in Fig. 3b as a more general adaptive paradigm.

Self-correction significantly decreased the error rate on the first iteration, and slightly on the next few, as shown in Section 4. The success of this bootstrapping scheme prompted Henry Baird to try to replicate it on his celebrated 100-font dataset (by far the largest at the time). [54]. "Self-corrective" was recently revived for camera-based scene-text labeling [55,56].

Fig. 4 is a graphic illustration of why the multi-font class centroids tend to move towards the single-font centroids in spite of misclassified samples.

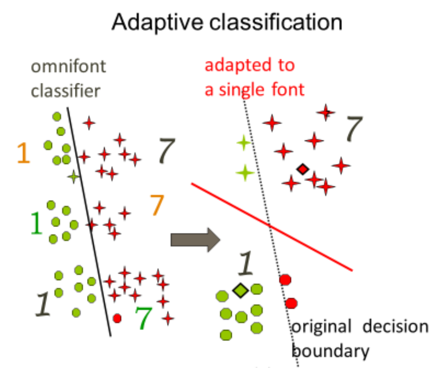


Fig. 4. A two-class ("1" vs. "7") classifier where the classification boundary is the hyperplane that bisects the line segment that connects the class centroids. There are three fonts: Calibri, Bookman, and Lucida, with the samples of each shown in different colors. The red solid line is on the right is the hyperplane r defined by the centroids (shown as diamonds) of the Lucida samples classified as "1" or "7" by the omnifont classifier. The omnifont classifier makes 3 errors, but the retrained final classifier is 100% correct.

Oddly enough, recycling only samples classified with high confidence, which occurs as a possible improvement to everyone who considers this scheme, usually yields *worse* results. My conjecture is that it is precisely the patterns nearest the classification boundaries that do most of the work (as suggested by the active learning paradigms of § 2.3).

Ho and Agrawala offered a theoretical perspective on self-correction soon after we published our first results [57,58]. We still do not know, however, why simple mean adaptation works as well as it does, and when it is likely to fail. While developing sufficient (albeit unrealistic) conditions for successful adaptation is easy, after fifty years we still cannot formulate *necessary* conditions. Finding necessary conditions for the more complex self-learning methods will be even more difficult. The empirically observed magic of adaptive retraining of a simple classifier by an error-prone omnifont OCR engine is cogently discussed by industrial experts Ray Smith in [59] and István Marosi in [60] (both also emphasize the crucial importance of meticulous layout analysis).

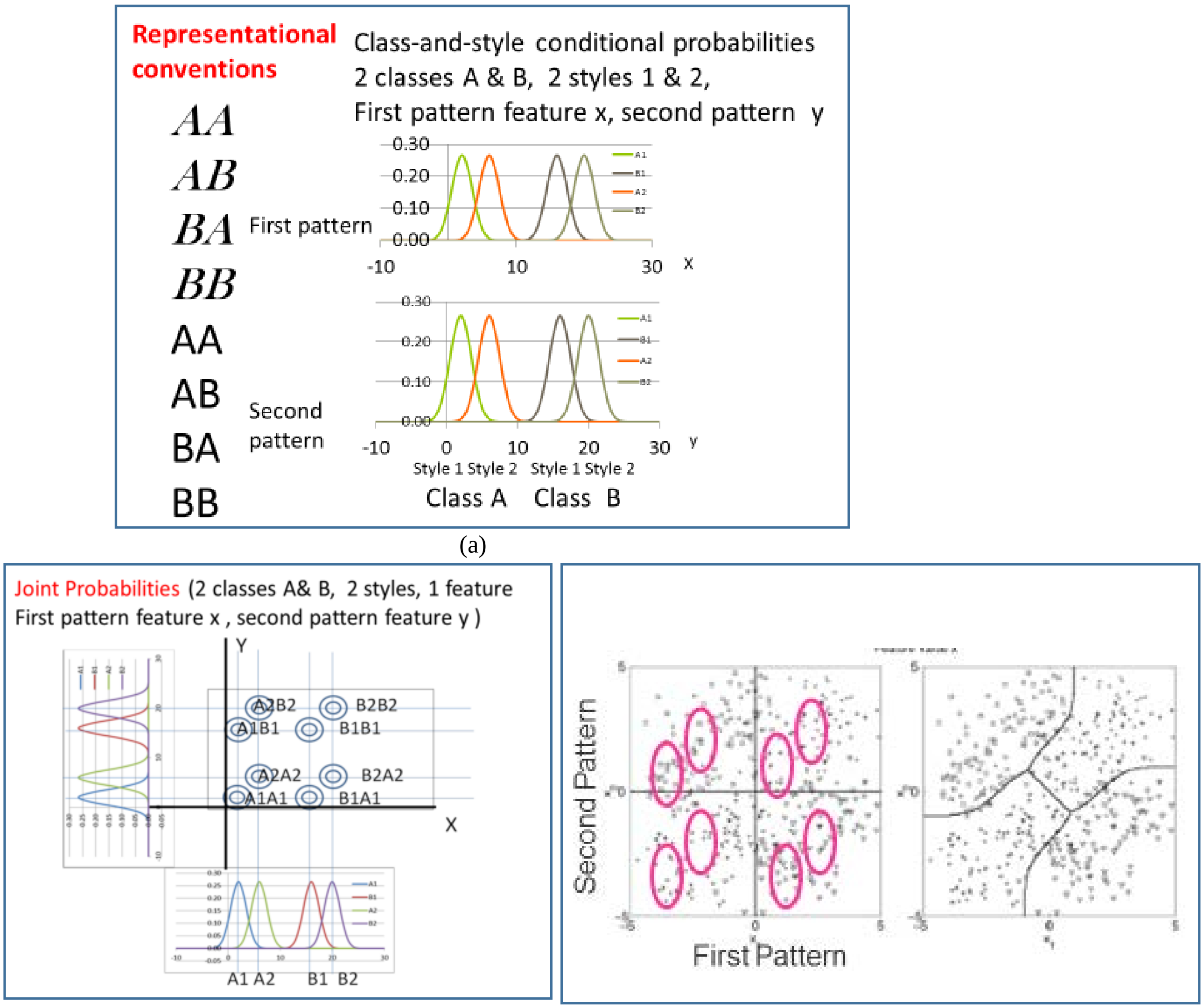
Eventually Harsha Veeramachaneni improved on self-correction by regularizing the covariance matrices of the training set and by adapting the covariance matrices as well as the mean vectors via Expectation Maximization [61]. The largest improvements over the static Bayesian classifier were due to regularization (necessitated by too few samples) and mean adaptation, but re-estimating the covariance matrices also contributed.

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

3.2 Style constraints

Style consistency is what Prateek Sarkar called the statistical dependence between the feature vectors of same-source (*isogenous*) samples. He defined *discrete* and *continuous* styles and *weak* (intra-class) and *strong* (inter-class) style consistency, and devised and tested optimal and fast suboptimal (style-consistent / -conscious / -constrained) classifiers [62]. Sarkar also developed the diagrams of Fig 5 that we later used to illustrate other style-conscious classification algorithms.

Both adaptive and style-constrained classifiers are field classifiers as opposed to singlet classifiers. That means that the classification of every pattern in the field benefits from information from every other pattern. In our terminology, the field of an *adaptive* classifier encompasses the entire test set, while *style-constrained* fields are of fixed length. The effects of this distinction on interaction between training and testing, approximate vs. optimal algorithms, classifier design, and the contrasting aspects of language-context based classification [63,64] and of font recognition [65,66,67], are elaborated with a highlight on context in [68] and on interaction in [69]. Now that is a non-explanation with abundant references!



(a) (b) (c)

Fig. 5. Pairs from two classes *A* and *B*, and two styles *1*, and *2*. (a) Individual Gaussian class-conditional probabilities (b) Joint probability contours (the samples of each pair have the same style) (c) Classification boundaries without and with style constraints.

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

Harsha Veeramachaneni liked to say that "the way Ann writes a *1* suggests how she would write a *7*". The key is the consistent *inter-pattern f eature dependence* that is neglected by conventional classifiers (Fig. 6). Style-consistency differs from language context that is usually defined in the OCR literature as statistical dependence between *labels* of the samples.

Fig. 6. Style. The *7* on the left is identical to the *1* on the right, but we can recognize them because the left *1* is different from the right *7*.

Harsha formulated adaptive quadratic discriminants for continuous Gaussian styles. He proved that under realistic assumptions, the covariance matrix of a same-source sequence of any length can be derived from the class-pair covariance matrices (though even for only ten classes and 100 features, 55 100x100 matrices must be estimated for each style!). He showed that as the same-source test fields get longer, the error rate converges to that of the intra-style (single-source) classifier [70] and derived sensible bounds on the reduction of the error rate with field length. He pointed out that field classifiers can be optimized either for minimum field error rate (as would seem desirable for short fields like zip codes), or for singlet error rate (for longer fields like business letters).

In a more abstract vein, Harsha proved that, style context, unlike language context, must be *order-independent*, e.g., in Duda&Hart notation:

$$P(5\ 7\ 9 \mid [\, \mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3]) = P(7\ \ 5\ \ 9 \mid [\, \mathbf{x}_2\ \ \mathbf{x}_1\ \ \mathbf{x}_3],$$

which is the *exchangeability* property that underlies several useful results of probability theory [71]. This property implies that style and language context require different methods, although, they can be combined for mutual benefit.

Style context is by no means a rare or an esoteric phenomenon in digitized documents (or in other classification tasks like faces or works of art). Aside from the obvious consistency of typeface in printed passages and the individuality of handwriting, more subtle effects arise from commonality of printer, pen and paper, and in-house standards of form and table layout. Indeed, one has to work hard to remove all style context (as perhaps desirable only in a ransom note). It has been clear for some time that further improvement in the accuracy of commercial OCR and off-line hand-writing recognition is unlikely without incorporating more context of every kind [72]. Research continues apace: a recent successful idea is Zhang's and Liu's *style transfer mapping* [73].

As indicated in§5.1.1.2.1 *Human-Machine Interaction Testbeds* and in §5.1.1.2.2 *Life-Long and Life-Wide Personal Assistant Platforms* of the influential *A 20-Year Community Roadmap for Artificial Intelligence Research in the US* [74], trends are in AI research are veering from transcription and labeling to content understanding, high-level human-computer interaction, and natural language translation.

4. Illustrations from our Experiments

This section illustrates, in approximate chronological order, some of our experimentation with interaction and adaptation over the last half century. Light reading, I promise!

Typewritten characters. Fig. 7 reproduces the twelve impact-printer fonts and the corresponding error rate vs. iteration plots of the 1966 self-corrective classifier study described in Section 3.

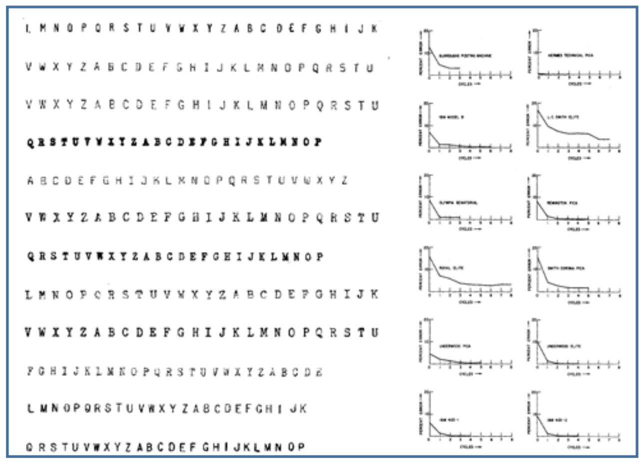


Fig. 7. Self-corrective character recognition. Fonts on the left, error rates vs. iteration on the right. The top of the y-axis is 20% [53].

Although OCR systems no longer attempt to recognize characters one at a time, adaptation to imperfectly machine-labeled text may reduce the voracious appetite for training samples of window-based systems like Hidden Markov Models and Convolutional Networks [75].

Printed pages. My first interaction with a scanned typeset (rather than typewritten) document was circa 1969. Fig. 8 shows field definition for line finding and character acquisition on a cathode ray tube attached to the scanner and to a graphics tablet, and labeling on the IBM 2250. I retrieved and retained the original figure captions from [4] and apologize for the quality of these fifty-year old Polaroid photos of an asynchronous CRT vector monitor. They illustrate tools for human-computer interaction before the advent of raster graphics and touch screens.

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

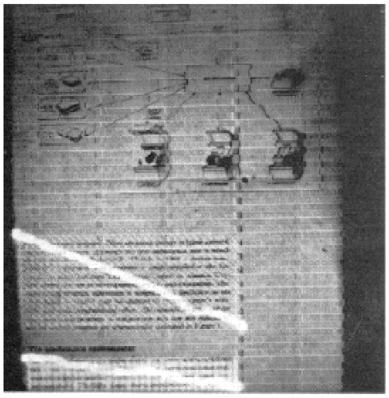
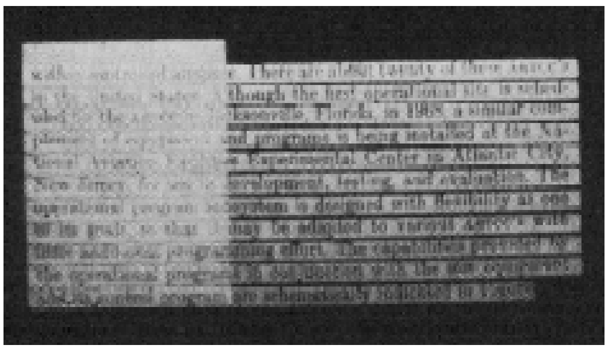
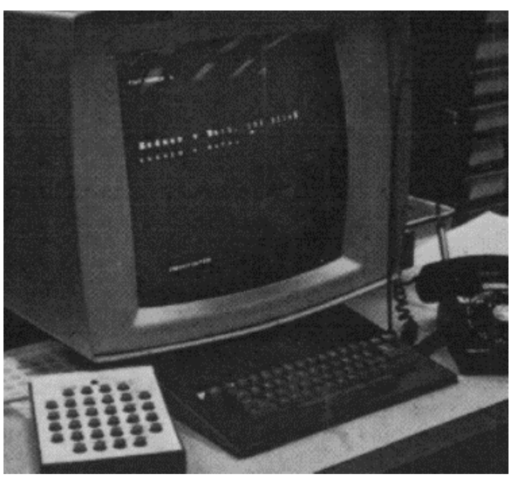


Fig. 8. (a) Field definition. The trace of the stylus on the display. In this time exposure one may also see the rectangular overlay which helps the operator to confirm that the boundaries have been correctly demarcated. In this example, two text fields have been selected.



(b) Character acquisition. The brighter region on the left is created by a time exposure of the horizontal line-finding scan. The character acquisition scan, consisting of short vertical strokes, is initiated as soon as a line is found. The shutter was closed just before the process reached the end of the upper text field shown in Fig. 8a.



(c) Display station for character labeling and identification. The 2250 display in the course of identification of video from legal text. The characters on the bottom line have already been typed in by the operator. Part of the line is brighter because the shutter was not synchronized with the CRT sweep.

Soil maps and text editors. In Nebraska at the end of the seventies, we digitized soil maps [76]. It was interactive in the sense that the computer warned the operator about anomalies like leaking region boundaries or offsets at the edges of the sheet (Fig. 9). The reading required to familiarize ourselves with the many national projects aiming to computerize maps and other sources of geographic data led to our review of GIS (then called *Geographic Data Processing* or *GDP*), in *Computing Surveys* [77]. A year later we reviewed –also in *Computing Surveys* –interactions with text [78]. Your word-processor may be dumb, but it is much smarter than it used to be!

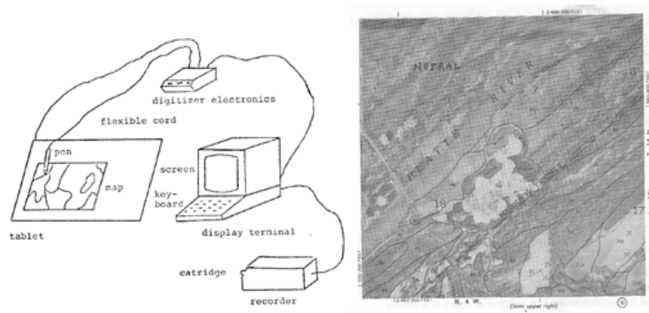


Fig. 9. Soil map digitization. Left: equipment. Right: a soil map.

Topographic maps. We next digitized street lines and recognized street names on a 1:24,000 USGS 7 ½ minute quad of Washington (D.C.) East (Fig. 10). After deleting contour lines and other map symbols via color and connected component analysis, the program reconnected broken street names, rotated them to horizontal, and extracted character prototypes to be used for classification prior to interactive correction of rejects [79,80]. Since all new maps are generated by computer, map conversion has become part of historical document processing. Nevertheless the extraction of map text that overlaps other symbology is still a topic of current research [81]. If others had not improved on our methods, mainly via more efficient interactive image processing, we would still drive around folding and unfolding road maps.

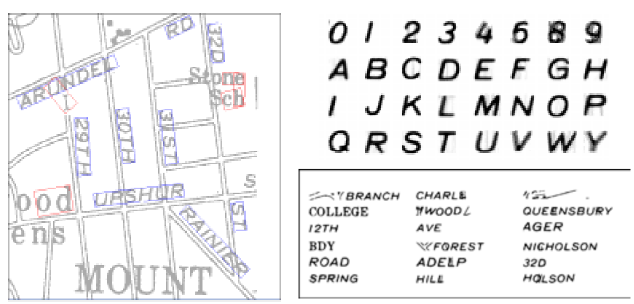


Fig. 10. Small part of a city map. Extracted street layer, character templates, and segmented street names grouped and rotated to horizontal for ease of automatic or manual transcription.

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

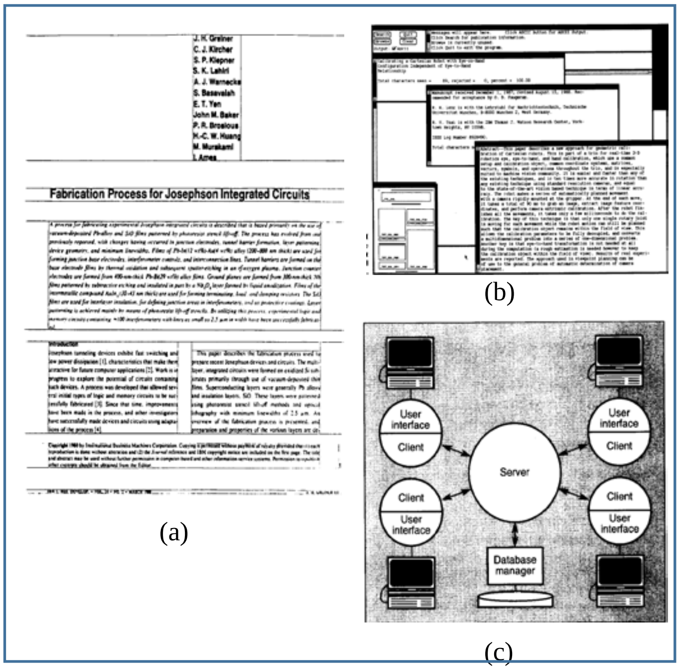


Fig. 11. Layout analysis. (a) X-Y tree for a title page from the IBM Journal; (b) text blocks from a PAMI page processed by a commercial OCR system; (c) Naïve, pre-web idea of a browser.

Page layout segmentation. Maps and computational geometry were great fun, but RPI did not have a geography department, so at the end of the eighties I returned to printed documents. We segmented scanned technical articles using X-Y trees [82] and page grammars into a dozen components *like title*, *author*, *footnote*, *reference*, and *photo* [83,84]. We proposed but never completed a browser for image and text segments (Fig. 11) that in any case would soon have been rendered obsolete by the advent of Netscape *Mosaic*.

Reprise: Self-corrective. Fig.12 shows Henry Baird's results on 6,400,000 characters (100-fonts with 80 symbols from each, constructed with his defect model). His conclusion: *good investment: large potential for gain, low downside risk*.

Size (pt)	Error reduction	% fonts improved	Best	Worst
6	x 1.4	100	x 4	x 1.0
10	x 2.5	93	x 11	x 0.8
12	x 4.4	98	x 34	x 0.9
16	x 7.2	98	x 141	x0.8

Fig. 12. Results of self-corrective recognition on Baird's database.

Style constraints. Weak style means that each class within a field, appears in a single style. Strong style implies that every sample within a field has the same style (Fig.13). Both occur in the wild.

	INTRA-CLASS	INTER-CLASS
Source 1:	22/07/1925	25/07/1922
Source 2:	25/05/1935	05/05/1925
Source 3:	21/06/1943	02/06/1943
Source 4:	03/24/1945	02/25/1942

Fig. 13. Two font variants in different configurations.

Adaptation. Fig. 14 shows results on an NIST dataset [61].

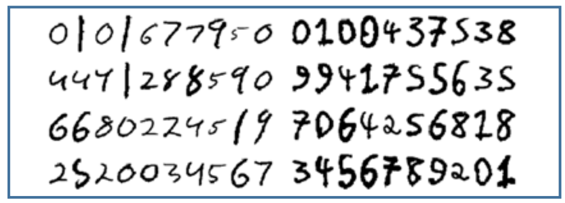


Fig. 14. Adaptive classification. Left: writers with increased errors (≥2 errors); Right: writers with fewer errors (≤26 errors).

Style-constrained classification. The benefit of style-constrained classification over singlet classification grows (slowly) with field length, as shown in Fig. 15 [70].

	Field error rate (%)			
Field length:	L=2		L=5	
Test data	w/o style	with style	w/o style	with style
SD3	1.4	1.3	3.0	2.5
SD7	7.4	6.7	13.8	11.7

Fig. 15. Field error rates on two NIST datasets, trained on ~ 17,000 characters and tested on ~17,000 characters with 5 top principal component "Hitachi" blurred directional features.

Tight print. Yihong Xu extracted character prototypes from homogenous page segments and devised a level-building algorithm to match them in various sequences to touching printed characters (Fig. 16) [85]. Crowded print is common in historical documents and in analog photocopies.

Tables. VeriClick is an interactive point-and-click tool for entering the four critical cells that suffice to segment the stub header, column header, row header and data regions of a table (Fig. 17). The tool may be primed with a file of critical cells generated by an imperfect segmentation program [86].

Computer Assisted Visual Interactive Recognition. CAVIAR is perhaps our best example of green interaction because it allow operator intervention to improve both feature extraction (the visual flower model) and top-ranked class assignments (into species) [87]. Fig. 18a shows the

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

GUI and Fig. 18b is a simplified diagram of the dataflow. We experimented with 1200 wildflower photographs of about 100 species that we collected ourselves (most enjoyable!), and with 400 NIST face images. We ported CAVIAR to two hand-held microprocessors with a camera [88]. We are often asked if CAVIAR could recognize birds, but that is a stretch. For actual use, any flower recognition system should incorporate the approximate dates when species bloom in various regions, and a GIS to provide latitude, elevation, and exposure (Easterly, Southerly). Several thousand species of wildflowers have been catalogued. CAVIAR needs at least one photo in situ (rather than a drawing from a flower guide) of each species. Perhaps they can be collected by bee-drones.

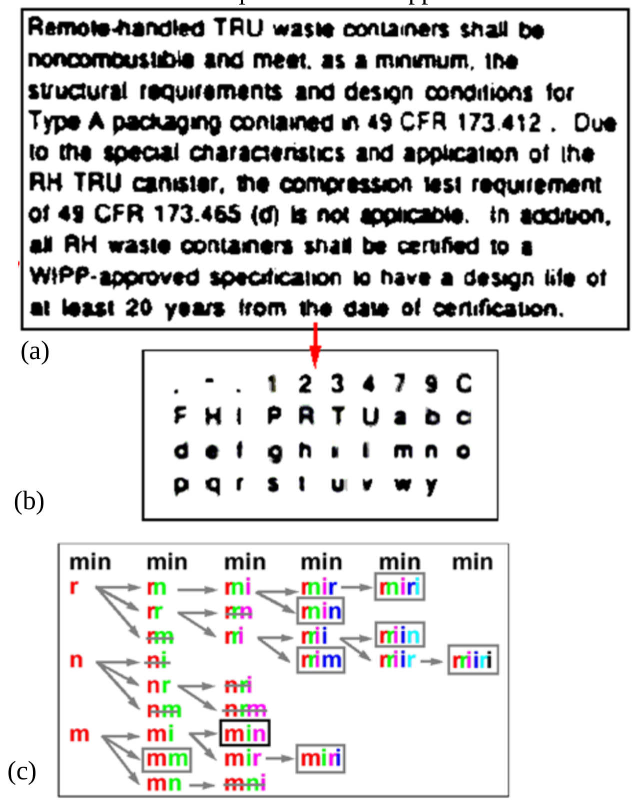
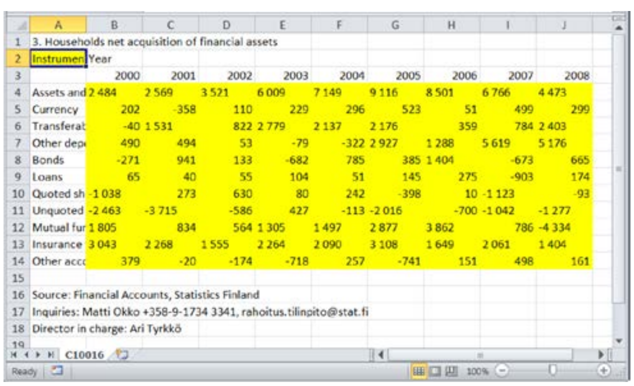


Fig. 16. Adaptive prototype extraction. (a) Ugly scanned text; (b) Extracted prototypes; (c) Level building: the correct solution, which best fits the input, emerges in the third row from the bottom.



	A	B	C	D	E	F	G	H	I	J
1	3. Households net acquisition of financial assets									
2	Instrumen	Year								
3		2000	2001	2002	2003	2004	2005	2006	2007	2008
4	Assets and	2 484	2 569	3 521	6 009	7 149	9 116	8 501	6 766	4 473
5	Currency	202	-358	110	229	296	523	51	499	299
6	Transferat	-40	1 531	822	2 779	2 137	2 176	359	784	2 403
7	Other dep	490	494	53	-79	-322	2 927	1 288	5 619	5 176
8	Bonds	-271	941	133	-682	785	385	1 404	-673	665
9	Loans	65	40	55	104	51	145	275	-903	174
10	Quoted sh	-1 038	273	630	80	242	-398	10	-1 123	-93
11	Unquoted	-2 463	-3 715	-586	427	-113	-2 016	-700	-1 042	-1 277
12	Mutual fur	1 805	834	564	1 305	1 497	2 877	3 862	786	-4 334
13	Insurance	3 043	2 268	1 555	2 264	2 090	3 108	1 649	2 061	1 404
14	Other acc	379	-20	-174	-718	257	-741	151	498	161
15										
16	Source: Financial Accounts, Statistics Finland									
17	Inquiries: Matti Okko +358-9-1734 3341, rahoitus.tilinpito@stat.fi									
18	Director in charge: Ari Tyrkkö									

C10016
Ready

Fig. 17. VeriClick. If the computer-generated segmentation is incorrect, the operator clicks on the corners of the stub head or the data region (both yellow) to correct it

Manuscripts. Dan Lopresti developed an interface for entering metadata for historical documents (Fig. 19). We explored possible paths from interaction to automation [89].

Survey forms and election ballots. Lopresti's GUI was designed for flexibility. He eventually modified it for several other applications, including the survey- and ballot-definition interface of Fig. 20 [90]. Endowing such tools-with green interaction requires combining them with automated processing to fill out the forms for operator correction of incorrect entries. We eventually programmed automated ballot tallies [91], but never got around to combining them with the GUI.

Calligraphy. Xiafen Zhang's Web interface (Fig.21) speeds up labeling databases of ancient calligraphy [92]. The pages are segmented semi-automatically, with interactive correction where necessary. Displaying each unknown sample with its neighbors helps to identify rare and ancient characters. As in CAVIAR, a classifier provides top candidates. Clicking on one of the candidates is faster than the default entry method via *Pinyin*. Operator actions are timed and logged for related human factors research.

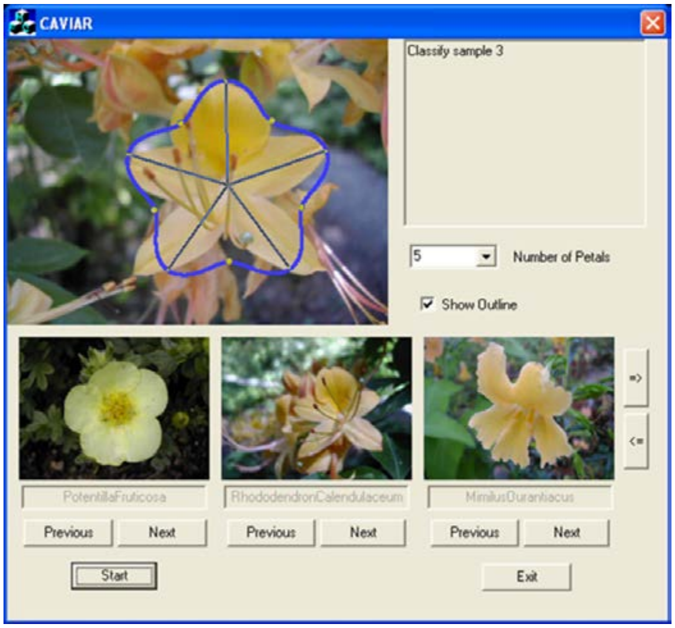
All the interfaces of Figures 17 – 21, are designed to display computer-generated labels and images that can be confirmed or corrected with a few clicks. Such classifier-assisted interactions are likely candidates for crowd-sourced labeling that requires minimal key entry.

Green information extraction from semi-structured books. At the suggestion of my decades-long collaborator David Embley, in 2016 I began development of multi-resolution template matching for extracting factoids (names, places and dates) from family books. These fragments of information are usually harvested from printed sources by volunteers. They are eventually integrated by FamilySearch, Ancestry and others into genealogical data bases.

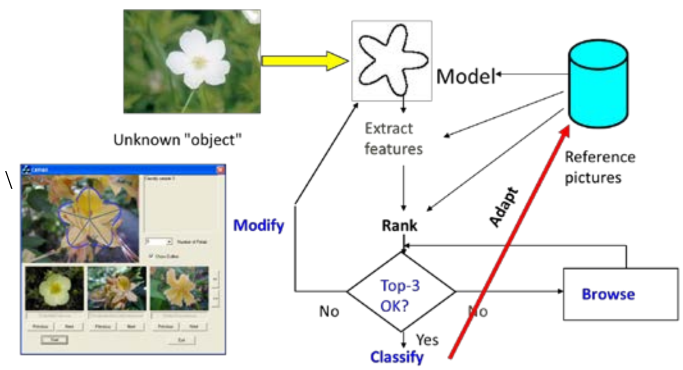
Searching family trees became a growth industry after their migration from microfilm to the web. A FamilySearch website mentions 300,000 scanned and OCR'd family books awaiting further processing.

It is easy enough to construct a dozen or so templates that will extract 90% or even 95% of the factoids from a book. This is, however, a long-tailed process: 98% may need ten times more user interaction, and 99% another factor of ten. It is increasingly hard for a volunteer to spot unusual text configurations (e.g. *natural* instead of *born*) in hundreds of very dry pages

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x



(a)



(b)

Fig. 18. CAVIAR (a) The GUI allows one to drag control points on the flower outline, to select one of the candidates as representing the species of the unknown, or to scroll through additional candidates. (b) When the visual model is changed, the ranking is recomputed. The classifier improves as new samples are identified and added.

The *green* aspect of our procedure is that after a first pass, with 90-95 of the factoids already labeled (mostly correctly), the program suggests locations in the text that are likely to yield good query phrases for new templates. The first results, based on a Scottish parish registry, Ohio funeral parlor records, and the Ely Family history (in toto ~250,000 factoids), were presented at a workshop last year [93]. I am almost ready to report further progress. I also eager to explain why, in spite of the spectacular results of deep learning on document and speech recognition, we chose a configuration of conventional NLP tools that allows fast and effective user interaction.

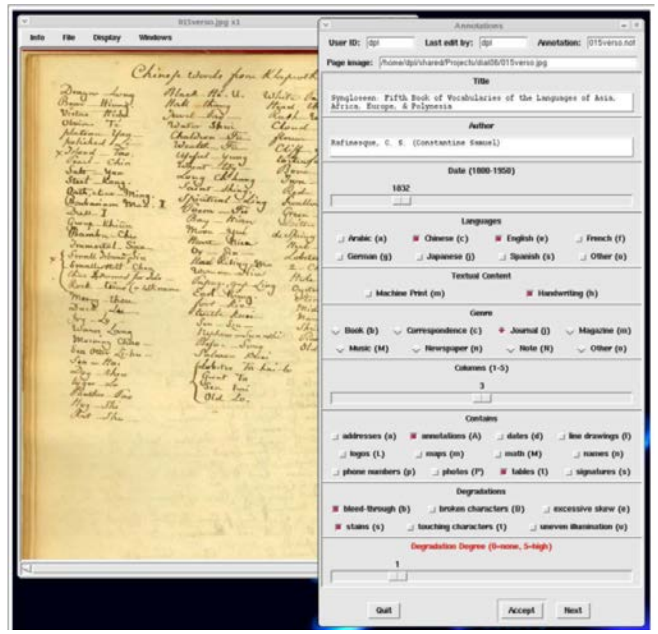


Fig. 19. Clicker, a prototype for collecting metadata.

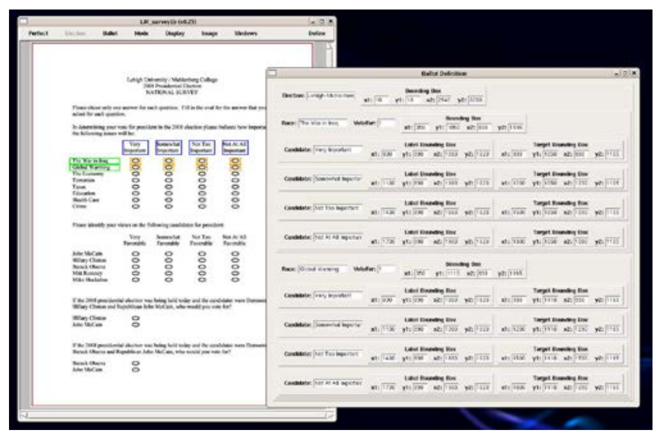


Fig. 20. Ballot Tool for priming automated processing.

Fig. 21. Web interface for computer-assisted labeling.

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

This was more than enough reminiscing! I enjoyed research on these topics (and on a few others), so much that I could not stop when I retired in 2011. It is now past time to assume a spectator's role, but first let me peculate a little about a recurring fantasy of mine, the Lifetime Reader [94].

5. The Lifetime Reader

The Lifetime Reader that I covet (*LiRe* for short, rhyming with *eyrie*) reads everything that I do, as well as most of what I see but skip. Unlike me, it remembers everything. It has no output other than a USB port or wireless, so I need a smartphone to retrieve anything. It is my personal web that differs from the worldwide web much as my backyard differs from a national forest. As the culmination of the efforts discussed in previous sections at constructing a universal reading machine, it offers exciting research opportunities on combining camera-based OCR with personal information retrieval. Version 1, the *memex*, was proposed by Vannevar Bush:in 1945: *As the scientist of the future moves about the laboratory or the field, every time he looks at something worthy of the record, he trips the shutter and in it goes, without even an audible click. Is this all fantastic*? [1].

5.1 Camera-based OCR

We have come a long way since my first exposure to camera-based OCR in September 1960 [95]. Fig.22 shows Mark 1 (now in the Smithsonian Museum) learning its letters.

During the Cold War, subminiature spy cameras captured readable document images. Millions of genealogical microfilm records are still being digitized inside Granite Mountain [96]. Research on camera-based OCR began long before the first workshop on Camera Based Document Analysis and Recognition (CBDAR) was held in conjunction with ICDAR 2005. At that ICDAR, Masashi Koga demonstrated mobile-phone Kanji OCR on snippets of printed pages [97]. Many of you will remember Liang's, Doermann's and Li's thoughtful survey in the same year [98]. Dave Doermann and his colleagues went on to rectify and mosaic camera-captured document images [99,100], paving the way for information extraction from full pages of text. Another approach is matching fragments of documents at the image level [101]. Instead of OCR-ing the page, it can be recognized and retrieved from a digital library [102,103].

The technology needed for LiRe is almost available and almost affordable. The wearable camera and, micro-processor can be far simpler than virtual reality headsets like Google Glass[104] or HoloLens[105] because all the downstream document processing, including rectification, layout analysis, OCR, perhaps translation, encryption, information retrieval, and display, will take place on another platform.

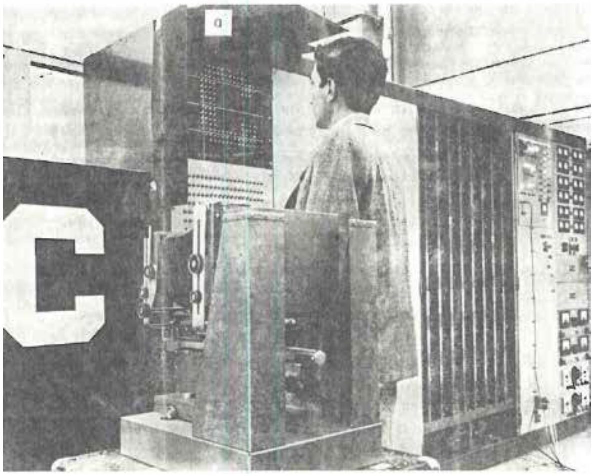


Fig. 23. Camera-based OCR in 1960 with a 20 x 20 pixel camera.

As expected with an idea whose time has come, other researchers have also proposed keeping track of what we read [106,107,108] or, as suggested much earlier, translating it into our own language [109].

5.2 Reading for Fun and Profit

What have I read or seen but remember only dimly if at all? I struggled through much of Kramèr's *Mathematical Methods of Statistics*, where some paragraphs took me an hour. I read *War and Peace* when it was obligatory, and a good deal of light fiction (some as e-books) that was not. I leafed through many pages but read only a few stories in thousands of issues of the *New York Times* and the Albany *Times Union.* I try to keep up with the *IAPR* and the *TC-11 Newsletters*, and occasionally revisit your conference papers and articles (in print from my file cabinets or on the web) when I forget crucial details or, less often, your major point.

I skim advertising pamphlets for hearing aids (all the manufacturers must know that I am nearly deaf), colorful political flyers requesting my vote, notices from my credit card issuer either initiating or canceling rental car insurance, and Privacy Policies. I dig out old email and saved directions to an obscure trailhead. Every year I re-read the instructions to relight our furnace pilot and program the thermostat.

I do remember parts of our interesting conversation on the way to the *ICPR* banquet in The Hague, but your name escapes me. I would like to glance at the business card you gave me then or the conference badge you wore. (Apparently others need help too [110].) We talked about scanner calibration. Now where did I see that advertisement for universal calibration charts for wide-angle cameras?

LiRe will save me hours of search for that play where the old curmudgeon insists on noodles without salt. Try finding it on-line! It will keep me from misquoting Shakespeare or Knuth. Is the warrantee on my watch still valid? My personal

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

Wikipedia will let me pose as an authority on the history of skeletonizing and stroke thinning. *Wholly new forms of encyclopedias will appear, ready made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified* [1].

Other LiRe users may want to retrieve elusive blogs, newsfeeds, webinars, wikis, genealogical resources, historical accounts that they had enjoyed, letters they wish they had kept, and PDF or DOC files that they never transferred from a broken laptop.

So what do we need to construct a LiRe?

5.3 LiRe Parts List

We need only a tiny high-resolution camera, a fast on-board processor with plenty of primary and secondary storage, and a garden-variety smartphone or laptop to which we can upload images of what we have recently read. The host computer will cull useless or repetitive images that were not filtered out by the camera computer, and OCR and index the rest for eventual regurgitation.

Camera. The camera hound of the future wears on his forehead a lump a little larger than a walnut. It takes pictures 3 millimeters square, later to be projected or enlarged, which after all involves only a factor of 10 beyond present practice. [1] A 60° field of view, with autofocus from 25 cm to infinity, could capture newspaper pages. Text that I merely glance at rather than read can be mosaicked with methods developed for copying for a large document on a small copier [111]. I envision a frame rate of 1/s, or perhaps slightly higher to provide redundant images.

A CCD or CMOS sensor array of 5K by 4K pixels (20 MB per frame) should be adequate for letter-sized text images that occupy only less than one quarter of the field of view. Borescope and endoscope cameras are small enough but they are limited to about 2 Megapixels. The camera should be worn as close as possible to the eye (mounted on spectacles, on the ear like *LifeLogger* [112]), or a collar or a lapel). Several patents have already been filed for cameras in smart contact lenses.

On-board Memory. The entire material of the Britannica in reduced microfilm form would go on a sheet eight and one-half by eleven inches [1]. A full-text page of *IJDAR* has about 1400 words. Perhaps some pages can be skimmed in a second or two, but at an average reading speed of 300 words/minute many pages will take 5 minutes. So for attentive reading retaining one frame per minute, rather than one per second, would suffice. We dwell on street signs, invoices, and most ads only briefly, but they have less text. We can count on a lossless compression ratio of at least 40:1 on clean text images scanned at 300 dpi [113]. So 17GB will suffice for the compressed images corresponding to turning a page every second for eight hours.

Speech and image recognition. A mic could add considerable functionality. Hearing aids already attempt to differentiate and filter intelligible speech from noise. Some of the processing software could be combined: OCR with a speech recognition toolkit was presented at a recent conference [114]. On the retrieval end, most of my friends prefer listening to reading while driving or exercising. This does not require additional provisions because audio replay and text-to-speech are already part of mobile platforms.

In principle, video can also be continuously collected, as famously suggested by Gordon Bell.. However, the plethora of tools for organizing personal photos suggests that finding desired pictures in even relatively small and purposive collections is far from easy. LiRe is only for text.

On-board microprocessor. The main task of the camera-integrated computer is to detect text and to compress and temporarily store the text images. Fast image and video compression algorithms are already available. Furthermore, not all the compression must be done in real time. Stored images could be filtered and compressed when the camera has no readable text in view. An alternative is to maintain the camera in a low-resolution "surveillance mode" except when it sees readable text.

Rapid but only partial processing of the data stream requires some research because most current camera-based OCR systems are designed for near-real-time output [115], as required, for example, for translating posted signs [116].

Off-board processing. The off-board computer has several tasks. It must combine some frames and discard others, perform layout analysis on a variety of input (text documents, forms, computer displays, scene text), carry-out OCR, perhaps translate and encrypt the transcript, update the index, accept user directions, and provide a query interface.

Often-noted differences between scanned and camera captured text are the possibility of severe geometric – affine and perspective – distortion, and contrast variations due to uncontrolled illumination. Although dozens of binarization and skew detection/removal methods are available, the extent of distortion in camera captured text, especially scene text, requires methods similar to those used in computer vision [117]. We expect, however, that the bulk of the text acquired by LiRe will have been purposively read and subject only to modest distortion. Most of us prefer to read in good light, and tend to keep reading material horizontal and perpendicular to our (and the camera's) line of sight A calculation in [94] shows that 20GB will suffice for a lifetime of *encoded* visible text and transcribed speech.

Retrieval. How can we find what we want in all that stored text? Ay, there is the rub. Query formulation for a LiRe personal collection will require new thought. Useful notions for personal filing systems were proposed and demonstrated

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

by Fujisawa et al. thirty years ago [118]. Although data structures for concept relations have evolved, such a system would still require user annotation at the front end (for which voice input would often be convenient). This would seem appropriate only for selected "important" text images. LiRe must also be able to collect text without distracting the user while she is reading, surfing the web, watching a presentation, or window–shopping. Fortunately complete temporal (calendar and clock) and spatial (GPS) referencing requires only existing technology.

For retrieval of non-annotated material, we could perhaps adopt and adapt browser technologies which are now well beyond simple keyword search. Initially there will not be any PageRank factor, but I do tend to look for the same thing more than once. Another set of query tools is available from the Library side, which started out with *Author* and *Subject* card catalogs but now incorporate all the tools of information retrieval like pattern matching on compressed text, inverted indices, perfect hashing, signature files, elaborate text tagging, fuzzy clustering, vector-space models, latent semantic indexing, graph algorithms, and relevance feedback. It is all turning into AI!

Three factors work in our favor. The first advantage over web search is that we are not talking about that much data to be indexed. Even if you acquired a LiRe in grade school and live to be a hundred, the final volume of 20 GB to be searched is less than one fifth of one millionth of the estimated size of the pages indexed by Google Search (100,000 TB in 2019). The second advantage is that the list of top-ranking items displayed in response to a query will already seem familiar, so we can parse it quickly to find the page, passage or phrase that we sought. (This is why some of us hang on to our obsolete but well-thumbed textbooks.) Finally, we will not be too bothered by OCR errors because we are all used to fractured and misspelled prose and because we are likely to keep what LiRe tells us to ourselves

While AI is heading towards content understanding and retrieval, we will looking only for what we might have read.

5.4 A Bucket List

In spite of the popularity of the biennial CBDAR and early successes on small-format input like business cards and on scene text, I have not yet seen any wearable-camera based system capable of reading magazine or book pages, let alone the New York Times. The LiRe is barely on the threshold of technological feasibility. Aside from configuring the required hardware, most of the research that remains to be done falls squarely in our community:

Text detection in spatial context, at home, at work, in local venues, in transit, abroad;

Mosaicking required by head and body motion;

Reading-order (no gaze tracking here);

Retention policy for unreadable and unindexable fragments of text, and near-duplicates;

Duplicate detection from consecutive frames and after (possibly lengthy) interruptions;

Lazy compression of sparse-text images;

Perspective-invariant recognition instead of rectification;

Adaptation to predictable reading material like the daily newspaper, magazines, the remaining volumes of the Jack Aubrey series, *IJDAR*, and *Python v2.7.6 documentation*;

Front-end annotation, for instance by tracing a phrase on a page or display with a deliberately selected finger or saying "at RPI Libe for preparing IJDAR paper.";

Back-end annotation, selective, quick and simple;

Retrieval strategies that mesh with our own mental recall;

Personalization to scripts and languages; reading speed; computer display settings; paper and laptop reading postures; work, leisure, shopping and napping habits;

Logging queries, responses, and user reactions for improving LiRe even as our own memory gets worse;

Security and privacy: what do these mean over a lifetime?

Lifetime text recording: What are the social, ethical,, legal and marketing implications?

This is surely an incomplete list, but there is already enough here to keep some graduate students busy. Some additional issues are raised in [9494]. I look forward to your and your students' (and perhaps to *their* students') contributions to the LiRe. If you are an entrepreneur, please remember that I am an eager potential customer.

5.5 End Note

I have been fortunate in having so many opportunities to learn from my own family, friends, colleagues and students. Thank you for bearing with this long-winded, green-tinted doc-tech auto-bio that lists my attempts to find new twists for automating symbol interpretation in the small and in the large. Documents are surely among the best examples of human achievement. That is why I continue to enjoy so much the quirky research problems, erratic progress, and excellent company of DA&R.

Acknowledgments

H. Fujisawa, P. Sarkar, A. Dengel, and three savvy IJDAR referees provided excellent suggestions. I am also grateful to the EICs of IJDAR, K. Kise, D. Lopresti and S. Marinai, who are (*disclosure*) old friends, for inviting me to ramble.

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

References

1 Vannevar Bush, As We May Think, The Atlantic, July 1945,

2 G. Nagy, A self-serving review of my own work, IAPR Newsletter, April 2012.

3 G. Nagy, Disruptive developments in document recognition, Pattern Recognition Letters, December 2015.

4 R.N. Ascher, G. Koppelman, M.J. Miller, and G. Nagy, "An Interactive System for Reading Unformatted Printed Text," IEEE Transactions on Computers, vol. 20, #12, pp. 1527-1543, December 1971.

5 R. G. Casey and G. Nagy, An autonomous reading machine. IEEE Trans. Computers, C-17(5): 492-503, May 1968.

6 R. G. Casey and G. Nagy, Advances in pattern recognition, Scientific American, 224(4): 56-71, 1971.

7 . Nagy, S. Seth, and K. Einspahr, Decoding substitution ciphers by means of word matching with application to OCR, IEEE Trans. Pattern Analysis and Machine Intelligence, 9(5): 710-715, September 1987.

8 T.K. Ho and G. Nagy, "OCR with no shape training," Proceedings of International Conference on Pattern Recognition-XV, vol.4, pp. 27-30, Barcelona, Spain, September 2000.

9 D. Blostein, G. Nagy, Asymptotic cost in document conversion, Procs. SPIE/EIT/DRR, San Francisco, Jan. 2012.

10 J. Zou and G. Nagy, Human-Computer Interaction for Complex Pattern Recognition Problems, in Data Complexity in Pattern Recognition (M. Basu & T.K. Ho, eds), Springer, London, 2006.

11 Y,T. Chien (1970) Interactive Pattern Recognition. Marcel Dekker Inc.

12 G.H. Ball, D.J. Hall, Some implications of interactive graphic computer systems for data analysis and statistics. Technometrics 12:17-31, 1970.

13 J.W. Sammon (1970) Interactive pattern analysis and classification. IEEE Trans. Computers 19:594-616

14 J. Tukey (1977) Exploratory data analysis, Addison-Wesley.

15 E. S. Gelsema, Applications of Interactive Pattern Recognition, in Pattern Recognition Theory and Applications, Proceedings of the NATO Advanced Study Institute, Oxford (Editors: J. Kittler, K-S Fu, L.F. Pau) 1981.

16 J.W. Smit, E.S. Gelsema, W, Huiges, R.F. Nawrath M.R. Halie MR, A commercially available interactive pattern recognition system for the characterization of blood cells: description of the system, extraction and evaluation of simple geometrical parameters of normal white cells, Clin Lab Haematol.; 1(2):109-119, 1979.

17 A.K. Jain and R.C. Dubes, Algorithms for Clustering Data, Prentice Hall, Englewood Cliffs, NJ 1988.

18 W. Siedlecki, K. Siedlecka, J. Sklansky (1988), An overview of mapping techniques for exploratory pattern analysis. Pattern Recognition 21:411-429.

19 J. Vesanto (1999) SOM-based data visualization methods. J. Intelligent Data Analysis 3:111-126.

20 T. K. Ho, Mirage: A Visual Tool for Scientific Inquiries, in M. Graham, M. Fitzpatrick, T. McGlynn (eds.), The National Virtual Observatory: Tools And Techniques For Astronomical Research, Astronomical Society of the Pacific, ASP Conference Series Vol. CS-382, 2008, 29-36.

21 G. Nagy, X. Zhang, Simple statistics for complex feature spaces, in Data Complexity in Pattern Recognition (M. Basu & T.K. Ho, eds), Springer, London, 2006.

22 E. A. Abbott, Flatland: A Romance of Many Dimensions, 1884.

23 G. Nagy, "Candide's Practical Principles of Experimental Pattern Recognition," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 5, #2, pp. 199-200, March 1983.

24 T. Novotny, Two challenges of correct validation in pattern recognition, Frontiers of Robotics and AI, Sep. 2014.

25 G. Nagy, "Document Image Analysis: Automated Performance Evaluation," Document Analysis Systems, A. L. Spitz and A. Dengel, Eds., pp. 137-156, World Scientific, Singapore, 1995.

26 S. Rice, G. Nagy, and T.A. Nartker, Optical Character Recognition: An Illustrated Guide to the Frontier, Kluwer Academic Publishers, Boston/Dordrecht/London, 1999.

27 T. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," Neural Computation 10, 1895-1923, 1998.

28 H.S. Baird, "Document Image Defect Models and Their Uses," Proc., IAPR 2nd Int'l Conf. on Document Analysis & Recognition, Tsukuba Science City, Japan, October 20-22, 1993.

29 Y. Li, D. Lopresti, G. Nagy, and A. Tomkins, "Validation of Image Defect Models for Optical Character Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, #2, pp. 99-108, February1996.

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

30 P. Sarkar, D. Lopresti, J. Zhou, and G. Nagy, "Spatial Sampling of Printed Patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, #3, pp. 344-351, March 1998.

31 I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, David; S. Ozair, C. A. Courville, Y. Bengio,. Generative Adversarial Networks. Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014). pp. 2672–2680.

32 J. Hu, R. Kashi D. Lopresti, G. Nagy, G. Wilfong, Why Table Ground-Truthing is Hard, Proceedings of the Sixth International Conference on Document Analysis and Recognition, September 2001, Seattle, WA, pp. 129-133.

33 D. Lopresti and G. Nagy, "Issues in Ground-Truthing Graphic Documents", Proceedings of the Fourth IAPR International Workshop on Graphics Recognition, September 2001, Kingston, Ontario, Canada, pp. 59-72.

34 B. Lamiroy and D. Lopresti, An Open Architecture for End-to-End Document Analysis Benchmarking, Procs. Int'l Conf. on Document Analysis and Recognition, 42-47, 2011.

35 D.D. Lewis, W.A. Gale, A sequential algorithm for training text classifiers, Procs. 11th Int'l COnf. on Machine Learning, 148-156, Morgan Kaufman, 1994.

36 G. Schohn, D. Cohn, Less is More: Active Learning with Support Vector Machines, Procs. Int'l Conf. on Machine Learning, 2000.

37 I., Dagan and S. Engelson, Committee-based sampling for training probabilistic classifiers, Int'l Conf. on Machine Learning, 1995.

38 S. Veeramachaneni, P. Avesani, Active sampling for feature selection, Procs Third IEEE International Conference on Data Mining, 665-668, 2003.

39 S Veeramachaneni, E Olivetti, P Avesani, Active sampling for detecting irrelevant features, Proceedings of the 23rd international conference on Machine learning, 961-968, 2006.

40 G. Nagy, Estimation, Learning, and Adaptation: Systems that Improve with Use, S+SSPR 2012, Pierre Devijver Award Lecture, Springer LNCS 7626, 1-12, 2012.

41 , L. Lipsky, D. Lopresti, and G. Nagy, Optimal Policy for Labeling Training Samples, Proceedings of Document Recognition and Retrieval XX (IS&T/SPIE International Symposium on Electronic Imaging), San Francisco, CA, pp. 865809-1 - 865809-9, 2013.

42 B. Gold, "Machine recognition of hand-sent Morse code," IRE Trans. Information Theory, vol. IT-5, pp. 17-24, March 1959.

43 D. B. Cooper and P. W. Cooper, "Nonsupervised adaptive signal detection and pattern recognition," D. B. Cooper and P. W. Cooper, Information and Control, vol. 7, pp. 416444, 1964.

44 S. C. Fralick, "The synthesis of machines which learn without a teacher," Tech. Rept. 6103-8, Stanford Electronics Lab., Stanford, Calif., April 1964; also "Learning to recognize patterns without a teacher," IEEE Trans. Information Theory, vol. IT-13, pp. 57-65, January 1967.

45 E. M. Braverman, "The method of potential functions in the problem of training machines to recognize patterns without a trainer," E. M.Braverman, Automation and Remote Control, vol. 27, pp. 1748-1771, October 1966.

46 A. A. Dorofeyuk, "Teaching algorithm for a pattern recognition machine without a teacher, based on the method of potential functions,", Automation and Remote Control, vol. 27, pp. 1728-1737, October 1966.

47 H. J. Scudder, III, "Probability of error of some adaptive pattern recognition machines," IEEE Trans. Information Theory, VOI. IT-11, pp. 363-371, July 1965.

48 R. W. Lucky, Automatic equalization for digital communication," Bell Sys. Tech. J., vol. 44, pp. 547-588, April 1965.

49 R. W. Lucky, "Techniques for adaptive equalization of digital communication systems," Bell Sys. Tech. J., vol. 45, pp. 255-286, February 1966.

50 J. Spragins, "Learning without a teacher," IEEE Trans. Information Theory, vol. IT-12, pp. 223-229, April 1966.

51 Ya. Z. Tsypkin (editor, Adaptation and learning in automatic systems, Academic Press, 1971.

52 Ya. Z. Tsypkin, Adaptation, training, and self-organization in automatic systems, Automation and Remote Control, vol. 27, pp. 1652, January 1966.

53 G. Nagy and G. L. Shelton, "Self-Corrective Character Recognition System," IEEE Transactions on Information Theory, vol. 12, #2, pp. 215-222, April 1966.

54 H.S. Baird and G. Nagy, "A Self-correcting 100-font Classifier," Proceedings of SPIE Conference on Document Recognition, vol.SPIE-2181, 106-115, San Jose, CA, 1994.

55 M. Tsukuda, M. Iwamura, K. Kise, Expanding Recognizable Distorted Characters Using Self-Corrective Recognition, 10th IAPR International Workshop on Document Analysis Systems (DAS), 327 – 332, 2012.

56 M. Iwamura, M. Tsukada, K. Kise, Automatic Labeling for Scene Text database, 12th International Conference on Document Analysis and Recognition (ICDAR), 1365 - 1369, 2013.

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

57 Y. C. Ho and A. K. Agrawala, "On the self-learning scheme of Nagy and Shelton," Proceedings of the IEEE, vol. 55, pp. 1764-1765, October 1967.

58 G. Nagy and N.G. Tuong, "On a Theoretical Pattern Recognition Model of Ho and Agrawala," Proceedings of the IEEE, vol. 56, #6, pp. 1108-1109, June 1968.

59 R. Smith, "An Overview of the Tesseract OCR engine," Procs. Document Analysis and Recognition ICDAR 2007, Curitiba, Brazil, 629-633, 2007.

60 I. Marosi, Industrial OCR approaches: architecture, algorithms and adaptation techniques, Procs. IS&T/SPIE Elecronic Imaging, DR&R. SPIE Vol. 6500, 1-10, 2007.

61 S. Veeramachaneni and G. Nagy, "Adaptive classifiers for multisource OCR," International Journal of Document Analysis and Recognition, vol. 6, #3, March 2004.

62 P. Sarkar and G. Nagy, "Style consistent classification of isogenous patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, #1, pp. 88-98, January 2005.

63 J. Raviv, Decision Making in Markov Chains Applied to the Problem of Pattern Recognition, IEEE Trans. Information Theory, VOL. IT-13, no. 4, 536-551, 1967.

64 G. T. Toussaint, The use of context in pattern recognition, Pattern Recognition, Vol. 10, 189-204, 1978.

65 A.W. Zramdini, R. Ingold, Optical Font Recognition Using Typographical Features, IEEE Trans. Pattern Analysis and Machine Intelligence, 20(8), 877-882, Aug. 1998.

66 H. Shi, T. Pavlidis, Font Recognition and Contextual Processing for More Accurate Text Recognition, Proc. Fourth Int'l Conf. Document Analysis and Recognition, vol. 1, pp. 39-44, 1997.

67 S. Marinai, E. Marino, E. G. Soda, Font adaptive word indexing of modern printed documents, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28 , 8, 1187 – 1199, Aug. 2006

68 S. Veeramachaneni, P. Sarkar, and G. Nagy, "Modeling Context as Statistical Dependence," Proceedings of Modeling and Using Context: 5th International and Interdisciplinary Conference CONTEXT 2005, Paris, France, Springer Lecture Notes in Computer Science, vol.3554, pp. 515-528, July 5-8, 2005.

69 G. Nagy, S. Veeramachaneni, "Adaptive and interactive approaches to document analysis," in Machine Learning in Document Analysis and Recognition (S. Marinai, H. Fujisawa, editors), Springer, Studies in Computational Intelligence, Vol. 90, pp. 221-257, 2008.

70 S. Veeramachaneni and G. Nagy, "Style context with second order statistics," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, #1, pp. 14-22, January 2005

71 S. Veeramachaneni and G. Nagy, "Analytical results on style-constrained Bayesian classification of pattern fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, #7, pp. 1280-1285, July 2007.

72 B. Klein, A. Dengel," Problem-adaptable document analysis and understanding for high-volume Applications," IJDAR 6, 3, 167-180, 2003.

73 X.Y. Zhang, and C-L Liu, Writer adaptation with style transfer mapping IEEE, Trans. Pattern Analysis and Machine Intelligence (PAMI) 35 (7), 1773-1787, 2013.

74 Computing Community Consortium and the Association for the Advancement of Artificial Intelligence, A 20-Year Community Roadmap for Artificial Intelligence Research in the US: https://cra.org/ccc/wp-content/uploads/sites/2/2019/08/17587_CCC-AI-Report_V7-1.pdf

75 B-L Cirstea, L. Likforman-Sulem, Improving a deep convolutional neural network architecture for character recognition, Proc. Document Recognition and Retrieval, IS&T Electronic Imaging, 1-7(7), February 2016.

76 G. Nagy and S. Wagle, "Approximation of Polygonal Thematic Maps by Cellular Maps," Communications of the ACM, vol. 22, #9, pp. 518-525, September 1979.

77 G. Nagy and S. Wagle, "Geographic Data Processing," ACM Computing Surveys, vol. 11, #2, pp. 139-181, June 1979.

78 G. Nagy and D.W. Embley, "Behavioral Aspects of Text Editors," ACM Computing Surveys, vol. 13, #1, pp. 33-70, March 1981.

79 G. Nagy, A. Samal, S. Seth, T. Fisher, E. Guthman, K. G. Nagy, A. Samal, S. Seth, T. Fisher, E. Guthman, K. Kalafala, L. Li, P. Sarkar, S. Sivasubramaniam, and Y. Xu, "A Prototype for Adaptive Association of Street Names with Streets on Maps," Graphics Recognition: Algorithms and Systems, K. Tombre and A. Chhabra, Eds., Springer Lecture Notes in Computer Science 1389, pp. 302-313, 1998.

80 L. Li, G. Nagy, A. Samal, S. Seth, and Y. Xu, "Cooperative text and line-art extraction from a topographic map," Proceedings of International Conference on Document Analysis and Recognition (ICDAR-99), pp. 467-470, Bangalore, India, September 1999.

81 N.H, Nazari,T. Tan, Tianxiang; Y-Y Chiang, Free Content Integrating Text Recognition for Overlapping Text Detection in Maps, Proc. Document Recognition and Retrieval, IS&T Electronic Imaging, 1-8(8), February 2016.

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

82 G. Nagy and S. Seth, "Hierarchical Image Representation with Application to Optically Scanned Documents," Proceedings of the Seventh International Conference on Pattern Recognition, pp. 347-349, Montreal, July 1984.

83 G. Nagy, S. Seth, and M. Viswanathan, "A Prototype Document Image Analysis System for Technical Journals," IEEE Computer, vol. 25, pp. 10-22, July 1992.

84 M. Krishnamoorthy, G. Nagy, S. Seth, and M. Viswanathan, "Syntactic Segmentation and Labeling of Digitized Pages from Technical Journals," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, #7, pp. 737-747, 1993.

85 Y. Xu and G. Nagy, "Prototype Extraction and Adaptive OCR," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, #12, pp. 1280-1296, December 1999.

86 G. Nagy, M. Tamhankar, VeriClick, an efficient tool for table format verification, Procs. SPIE/EIT/DRR, San Francisco, Jan. 2012.

87 J. Zou and G. Nagy, "Visible models for interactive pattern recognition," Pattern Recognition Letters Vol. 28, pp 2335-2342, 2007.

88 A. Evans, J. Sikorski, P. Thomas, S-H Cha, C. Tappert, G. Zou, A. Gattani, and G. Nagy, "Computer Assisted Visual Interactive Recognition (CAVIAR) Technology," 2005 IEEE International Conference on Electro-Information Technology, Lincoln, NE, May 2005 (Proceedings on CD-ROM only).

89 G. Nagy and D. Lopresti, "Interactive Document Processing and Digital Libraries," Procs. 2nd IEEE International Conference on Document Image Analysis for Libraries, Lyon, France, April 27-28, pp. 1-9, IEEE Computer Society Press 2006.

90 D. Lopresti, G. Nagy, E. Barney Smith, "A Document Analysis System for Supporting Electronic Voting Research," Procs. Document Analysis Systems, Nara, Japan, September 2008.

91 George Nagy, Daniel Lopresti, Elisa H. Barney Smith, Ziyan Wu, Characterizing Challenged Minnesota Ballots, Procs. Document Recognition and Retrieval, San Jose, SPIE, Jan 2011.

92 G. Nagy, X. Zhang, CalliGUI: Interactive Labeling of Calligraphic Character Images, Procs. ICDAR 11, Beijing, September 2011.

93 D.W. Embley and G. Nagy, Green Interaction for Extracting Family Information from OCR'd Books, Document Analysis Systems Workshop (DAS'18), Vienna, April 2018.

94 G. Nagy, The Lifetime Reader, IEEE Pervasive Computing,17, 4,86-95, 2018,

95 G. Nagy, "Neural Networks - Then and Now," IEEE Transactions on Neural Networks, vol. 2, #2, pp. 316-318, March 1991.

96 D. Ouimette, Digitizing the Records in the Granite Mountain, SamilySearch https://familysearch.org/learn/wiki/en/Digitizing_the_Records_in_the_Granite_Mountain accessed 4/11/2015.

97 M. Koga, R. Mine, T, Kameyama, T. Takahashi, M. Yamazaki, T. Yamaguchi, "Camera-based Kanji OCR for mobile-phones: practical issues," in Document Analysis and Recognition. Proceedings. Eighth International Conference 2, 29, 635-639, 2005

98 J. Liang, D. Doermann, H. Li, Camera-based analysis of text and documents: a survey, IJDAR 7:84,84-104,2005.

99 J. Liang, D. DeMenthon, D. Doermann, Geometric Rectification of Camera-Captured Document Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30, 4, 591 – 605, April 2008.

100 J. Liang, D. DeMenthon, D. Doermann, Mosaicing of camera-captured document images, Computer Vision and Image Understanding, Volume 113, Issue 4, April 2009, Pages 572–579

101 J. Moraleda and J.J. Hull, "Toward Massive Scalability in Image Matching," IAPR Int. Conf. on Pattern Recognition (ICPR), Istanbul, Turkey, Aug. 23-26, 2010, 3424-3427.

102 S. Ahmed, K. Kise, M. Iwamura, M. Liwicki, A. Dengel, Automatic Ground Truth Generation of Camera Captured Documents Using Document Image Retrieval. ICDAR 2013: 528-532

103 T. Nakai, K. Kise, M. Iwamura, Camera-based document image retrieval as voting for partial signatures of projective invariants, 8th International Conference on Document Analysis and Recognition (ICDAR), 379 - 383, 2005.

104 T. Toyama, A. Dengel, W. Suzuki, K. Kise, Wearable Reading Assist System: Augmented Reality Document Combining Document Retrieval and Eye Tracking, 12th International Conference Document Analysis and Recognition (ICDAR), 2013 on 30 – 34, 2013.

105 E.E. Sabelman, R. Lam, The real-life dangers of augmented reality, IEEE Spectrum 52, 7, 48 – 53, July 2015.

106 T. Kimura, R. Huang, S. Uchida, M. Iwamura, S. Omachi, K. Kise:The Reading-Life Log - Technologies to Recognize Texts That We Read. ICDAR 2013: 91-95

107 K. Kunze, K. Masai, M. Inami, Ö. Sacakli, M. Liwicki, A. Dengel, S. Ishimaru, K. Kise: Quantifying reading habits: counting how many words you read. UbiComp 2015: 87-96

Sept. 23, 2019 as Springer "Online First": Nagy, G. IJDAR (2019). https://doi.org/10.1007/s10032-019-00344-x

108 M. Matsubara, J. Folz, T. Toyama, M. Liwicki, A. Dengel, K. Kise: Extraction of read text using a wearable eye tracker for automatic video annotation. UbiComp/ISWC Adjunct 2015: 849-854

109 H. Fujisawa, H. Sako, Y. Okada, and S-W. Lee, "Information Capturing Camera and Developmental Issues," Proc. Int. Conf. Document Analysis and Recognition, ICDAR '99, Bangalore, India, Sep. 20-22, 1999, pp. 205-208.

110 M. Iwamura, K. Kunze, Y. Kato, Y. Utsumi, K. Kise: Haven't we met before? a realistic memory assistance system to remind you of the person in front of you. Augmented Human Int'l Conf. 2014: 32:1-32:4

111 J. Cullen and J.J. Hull, ''Oversize document copying system,'' IAPR Workshop on Document Analysis Systems, Malvern, Pennsylvania, October 14-16, 1996.

112 I. Frazer, Got a Bad Memory? This Company Has You Covered, Wall St Daily, Aug. 29, 2015.

113 R.N. Ascher and G. Nagy, "A Means for Achieving a High Degree of Compaction on Scan-Digitized Printed Text," IEEE Transactions on Computers, vol. 23, #11, pp. 1174-1179, October 1974.

114 P. Schone, A.Cannaday, S.Stewart, R; Day, J. Schone, Automatic Transcription of Historical Newsprint by Leveraging the Kaldi Speech Recognition Toolkit, Proc. Document Recognition and Retrieval, IS&T Electronic Imaging, 1-10(10), February 2016.

115 T. Kobayashi, M. Iwamura, T. Matsuda, K. Kise, An Anytime Algorithm for Camera-Based Character Recognition, 12th International Conference on Document Analysis and Recognition (ICDAR), n 1140 - 1144, 2013.

116 M. Mantha and J K. Chaithanya, Vision based Traffic Panel Text Information and Sign Retrieval, Int'l J. Current Engineering and Technology, Vol.5, No.4 (Aug 2015)

117 T. Nakai, K. Kise, M. Iwamura, Camera-based document image retrieval as voting for partial signatures of projective invariants, 8th International Conference on Document Analysis and Recognition (ICDAR), 379 - 383, 2005.

118 H. Fujisawa, A. Hatakeyama, and J. Higashino, "A Personal Universal Filing System Based on the Concept-Relation Model," Proc. 1st Int. Conf. Expert Database Systems, Charleston, SC, 1986, pp. 31-44.