DAS 94

International Association for Pattern Recognition Workshop on

DOCUMENT ANALYSIS SYSTEMS

Editors

A. Lawrence Spitz & Andreas Dengel

BIBLIOTHEQUE DU CERIST

SERIES IN MACHINE PERCEPTION & ARTIFICIAL INTELLIGENCE
Volume 14

World Scientific

Series in Machine Perception and Artificial Intelligence – Vol. 14

International Association for Pattern Recognition Workshop on

DOCUMENT ANALYSIS SYSTEMS

BIBLIOTHEQUE DU CERIST

Editors

A. Lawrence Spitz
Fuji Xerox Palo Alto Laboratory

Andreas Dengel
DFKI

World Scientific
Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 9128

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

BIBLIOTHEQUE DU CERIST

-1995

DOCUMENT ANALYSIS SYSTEMS
International Association for Pattern Recognition Workshop

Copyright © 1995 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 27 Congress Street, Salem, MA 01970, USA.

ISBN 981-02-2122-3

This book is printed on acid-free paper.

Printed in Singapore by Uto-Print

PREFACE

The first International Association for Pattern Recognition workshop on Document Analysis Systems was held in Kaiserslautern, Germany in October 1994. In planning this workshop, we were aware that the rapid growth of document analysis technology reflects considerable contemporary interest and anticipates large market potential and an important economic role in the coming decades. Therefore, we not only tried to attract academic and industrial researchers, but also vendors and end-users of document analysis technology. In total, we received 51 papers and 20 abstracts from 12 countries. The program committee was careful to uphold a high standard for contributions. We accepted 26 papers for oral presentation.

In assembling this book, we selected the papers most representative of progress in this field and augmented that selection with a few papers invited specifically for this volume.

The book is divided into sections about document analysis system architectures, classification and performance evaluation, handwriting and line-drawing recognition as well as innovative techniques and applications. We have also incorporated the reports of the eight working group sessions that took place during the Document Analysis Systems workshop.

We owe a special debt of gratitude to Henry Baird whose idea it was to hold a workshop on systems and who effectively and energetically promoted it. Finally, we would like to thank all authors for their contribution to this volume. We hope that this book will be useful and serve to promote further research in document analysis.

Palo Alto, California, USA — Larry Spitz
Kaiserslautern, Germany — Andreas Dengel

BIBLIOTHEQUE DU CERIST

BIBLIOTHEQUE DU CERIST

CONTENTS

System Architecture

Data Structures for Page Readers 3
H. S. Baird and D. J. Ittner

Palace: A Multilingual Document Recognition System 16
A. L. Spitz and M. Ozaki

Experiences with High-Volume, High Accuracy Document Capture 38
H. R. Stabler

OfficeMAID – A System for Office Mail Analysis, Interpretation and Delivery 52
A. Dengel, R. Bleisinger, F. Fein, R. Hoch, F. Hones and M. Malburg

Programmable Contextual Analysis 76
D. J. Ittner and H. S. Baird

A System for Exploiting Context in Automatic Recognition 93
R. A. Lorie

An Adaptive Approach to Document Classification and Understanding 114
S. W. Lam

Class Evaluation

Document Image Analysis: Automated Performance Evaluation 137
G. Nagy

Using Consensus Sequence Voting to Correct OCR Errors 157
D. Lopresti and J. Zhou

A Handwritten Character Recognition System by Efficient Combination of Multiple Classifiers 169
T. Suzuki, H. Nishida, Y. Nakajima, H. Yamagata, M. Tachikawa and G. Sato

BIBLIOTHEQUE DU CERIST

A Region-Based System for the Automatic Evaluation of Page Segmentation Algorithms 188
S. Randriamasy and L. Vincent

Integration of Contextual Knowledge Sources into a Blackboard-Based Text Recognition System 206
R. Sennhauser

Automatic Construction of Recognition Procedures for Musical Notes by Genetic Algorithm 225
I. Yoda, K. Yamamoto and H. Yamada

Recognition of Handwritten Responses on US Census Forms 237
T. M. Breuel

A System for the Recognition of Handwritten Literal Amounts of Checks 265
J.-C. Simon, O. Baret and N. Gorski

Handwritten Text Recognition 288
V. Govindaraju, R. K. Srihari and S. N. Srihari

Line Drawing

Knowledge Organization and Interpretation Process in Engineering Drawing Interpretation 307
P. Vaxiviere and K. Tombre

Processing Imprecise and Structural Distorted Line Drawings by An Adaptable Drawing Interpretation Kernal 318
B. Pasternak

Vector-Based Arc Segmentation in the Machine Drawing Understanding System Environment 338
D. Dori

Robust Drawing Recognition Based on Model-Guided Segmentation 363
S. Shimotsuji, O. Hori and M. Asano

Innovations

Document Image Matching and Retrieval with Multiple Distortion-Invariant Descriptors 379
J. J. Hull

BIBLIOTHEQUE DU CERIST

Off-Line Interpretation and Execution of Corrections on Text Documents 397
D. Möri and H. Bunke

Analysis of Scanned Braille Documents 413
R. T. Ritchings, A. Antonacopoulos and D. Drakopoulos

Document Analysis by Fractal Signatures 422
H. Ma, D.-H. Xi, X.-G. Mao, Y.-T. Tang and C.-Y. Suen

Working Groups

Possibilities for International Collaboration 445
A. Dengel and B. B. Jean

Document Analysis and Learning 450
J. Hull and T. Bayer

Needs of the Market and User Requirements 452
H. Stabler and H. Fujisawa

Evaluation–Criteria 455
H. S. Baird and J. Kanai

Handwriting 458
T. Breuel and J. M. Simon

Line Drawing and Music Recognition 462
S. Bauman and K. Tombre

Multilingual Documents and Natural Language Processing 465
M. Malburg and A. L. Spitz

Form Recognition 469
T. Caesar and R. G. Casey

BIBLIOTHEQUE DU CERIST