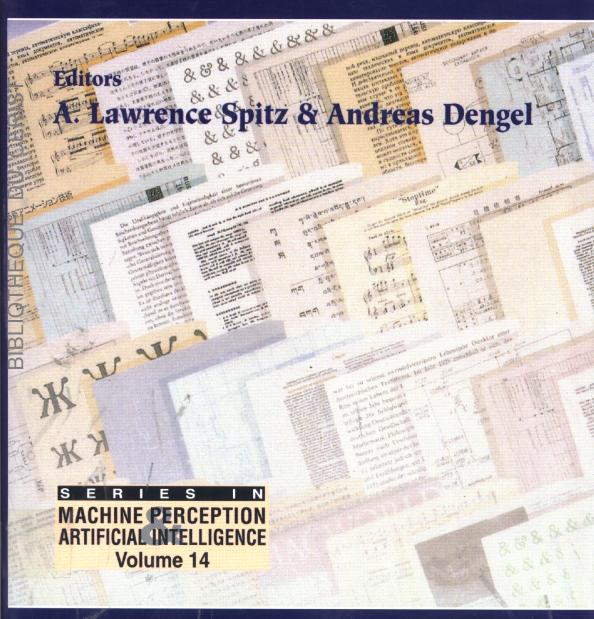


International Association for Pattern Recognition Workshop on

DOCUMENT ANALYSIS SYSTEMS



International Association for Pattern Recognition Workshop on

DOCUMENT ANALYSIS SYSTEMS

Editors

A. Lawrence Spitz

Fuji Xerox Palo Alto Laboratory

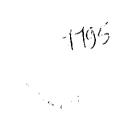
Andreas Dengel DFKI



Published by

World Scientific Publishing Co. Ptc. 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



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 Kaiserslautern, Germany Larry Spitz Andreas Dengel

CONTENTS

System Architecture	
Data Structures for Page Readers H. S. Baird and D. J. Ittner	3
Palace: A Multilingual Document Recognition System A. L. Spitz and M. Ozaki	16
Experiences with High-Volume, High Accuracy Document Capture H. R. Stabler	38
OfficeMAID – A System for Office Mail Analysis, Interpretation and Delivery A. Dengel, R. Bleisinger, F. Fein, R. Hoch, F. Hones and M. Malburg	52
Programmable Contextual Analysis D. J. Ittner and H. S. Baird	76
A System for Exploiting Context in Automatic Recognition R. A. Lorie	93
An Adaptive Approach to Document Classification and Understanding S. W. Lam	114
Class Evaluation	
Document Image Analysis: Automated Performance Evaluation G. Nagy	137
Using Consensus Sequence Voting to Correct OCR Errors D. Lopresti and J. Zhou	157
A Handwritten Character Recognition System by Efficient Combination of Multiple Classifiers T. Suzuki, H. Nishida, Y. Nakajima, H. Yamagata, M. Tachikawa	169

A Region-Based System for the Automatic Evaluation of Page Segmentation Algorithms S. Randriamasy and L. Vincent	188
Integration of Contextual Knowledge Sources into a Blackboard-Based Text Recognition System R. Sennhauser	206
Automatic Construction of Recognition Procedures for Musical Notes by Genetic Algorithm 1. Yoda, K. Yamamoto and H. Yamada	225
Recognition of Handwritten Responses on US Census Forms T. M. Breuel	237
A System for the Recognition of Handwritten Literal Amounts of Checks JC. Simon, O. Baret and N. Gorski	265
Handwritten Text Recognition V. Govindaraju, R. K. Srihari and S. N. Srihari	288
Line Drawing	
Knowledge Organization and Interpretation Process in Engineering Drawing Interpretation P. Vaxiviere and K. Tombre	307
Processing Imprecise and Structural Distorted Line Drawings by An Adaptable Drawing Interpretation Kernal B. Pasternak	318
Vector-Based Arc Segmentation in the Machine Drawing Understanding System Environment D. Dori	338
Robust Drawing Recognition Based on Model-Guided Segmentation S. Shimotsuji, O. Hori and M. Asano	363
Innovations	
Document Image Matching and Retrieval with Multiple Distortion-Invariant Descriptors J. J. Hull	379

Off-Line Interpretation and Execution of Corrections on Text Documents D. Möri and H. Bunke	397
Analysis of Scanned Braille Documents R. T. Ritchings, A. Antonacopoulos and D. Drakopoulos	413
Document Analysis by Fractal Signatures H. Ma, DH. Xi, XG. Mao, YT. Tang and CY. Suen	422
Working Groups	
Possibilities for International Collaboration A. Dengel and B. B. Jean	445
Document Analysis and Learning J. Hull and T. Bayer	450
Needs of the Market and User Requirements H. Stabler and H. Fujisawa	452
Evaluation-Criteria H. S. Baird and J. Kanai	455
Handwriting T. Breuel and J. M. Simon	458
Line Drawing and Music Recognition S. Bauman and K. Tombre	462
Multilingual Documents and Natural Language Processing M. Malburg and A. L. Spitz	465
Form Recognition T. Caesar and R. G. Casey	469