

Modeling for part-based visual object detection based on local features

**Von der Fakultät für Elektrotechnik und Informationstechnik
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines Doktors
der Ingenieurwissenschaften genehmigte Dissertation**

vorgelegt von

Diplom-Ingenieur
Mark Asbach
aus Neuss

Berichter:

Univ.-Prof. Dr.-Ing. Jens-Rainer Ohm
Univ.-Prof. Dr.-Ing. Til Aach

Tag der mündlichen Prüfung: 28. September 2011

Diese Dissertation ist auf den Internetseiten der
Hochschulbibliothek online verfügbar.

Aachen Series on Multimedia and Communications Engineering

Volume 8

Mark Asbach

**Modeling for part-based visual object detection
based on local features**

Shaker Verlag
Aachen 2011

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Zugl.: D 82 (Diss. RWTH Aachen University, 2011)

Copyright Shaker Verlag 2011

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-0581-3

ISSN 1614-7782

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen

Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9

Internet: www.shaker.de • e-mail: info@shaker.de

Dla Oli

Acknowledgements

Few dissertations have been written by eremites contemplating a scientific subject solely on their own, and certainly this book is not one of those rare exceptions. Over the last years I had the opportunity to meet great people with great ideas that profoundly influenced and sharpened my grasp of computer vision and machine learning problems. Moreover I have collected significant debts during the writing of this thesis, that I want to acknowledge here.

First of all, I would like to thank everyone at the Institute of Communications Engineering—and in particular my supervisor Prof. Dr.-Ing. Jens-Rainer Ohm—for the open, supportive and friendly atmosphere. You have made this work possible. Additional thanks go to the many friends who discussed the diverse aspects of my research with me, debated science and life in general, and who encouraged me to keep going all the way through. Finally, I want to express my gratitude to my colleagues at Fraunhofer IAIS for inspirational talks and discussions.

No matter how much effort put into it, this work will doubtlessly contain flaws, omissions and over-simplifications, that I take full responsibility for. Still, I hope that the presented thoughts and experiments might serve as a little piece of jigsaw in the understanding of object detection.

Contents

1	Introduction	1
1.1	The need for object detection	1
1.2	Motivation for part-based approaches	2
1.2.1	Effects of deformation and pose variation	2
1.2.2	Biological motivation	3
1.3	Research question	4
1.4	Contribution of this thesis	4
1.4.1	Semi-supervised identification of the object parts	5
1.4.2	Generic feature extraction	5
1.5	Organization of this thesis	6
2	State of the art	7
2.1	What is object detection?	7
2.2	Performance measures for object detection	7
2.3	Object detection algorithms	9
2.3.1	Segmentation-based approaches	10
2.3.2	The sliding window approach	11
2.3.3	Historical examples of the sliding window approach	12
2.3.4	Boosted cascade of simple features	12
2.3.5	Part-based approaches	13
2.4	Local image features	16
2.5	Performance evaluation of face detection algorithms	18
2.5.1	Evaluation data for part-based approaches	21
2.5.2	Databases used here	22
3	Part-based object detection using local image features	24
3.1	Image model	26
3.2	Object model	28
3.2.1	Local appearance classification	29
3.2.2	Generation of object hypotheses from parts	30
3.2.3	Geometry classification	33
4	Application to human faces	37
4.1	Dataset preparation	38
4.2	Deriving object parts	41
4.2.1	Sensitivity to object features	41
4.2.2	Distribution of feature locations	43
4.2.3	Identifying parts	45

4.3	Learning local appearance	48
4.3.1	Estimation of achievable classifier performance	50
4.3.2	Training setup for boosted cascades	52
4.3.3	Classifier performance on training set	54
4.4	Learning geometry	56
4.4.1	Finding the transformation matrix	56
4.4.2	Verifying the model on the training set	58
4.4.3	Finding optimum thresholds	59
4.5	Performance on training set	60
4.5.1	Number of parts per model	60
4.5.2	Number of Gaussian mixture components	61
4.5.3	Number of parts per constellation	61
5	Experimental validation	64
5.1	Results on the BioID database	65
5.2	Results on the AR face database	66
5.2.1	Results on artificially occluded faces	66
5.2.2	Results on faces occluded by sunglasses or scarfs	68
5.3	Results on the MPEG-7 VCE dataset	69
5.4	Influence of the error measure	69
5.5	Discussion	72
6	Conclusion	75
6.1	Summary	75
6.2	Conclusions	76
6.3	Outlook	77
A	Databases	79
A.1	AR face database	79
A.1.1	Downloaded archives	80
A.1.2	Duplicate images	80
A.1.3	Missing images	81
A.2	BioID database	81
A.2.1	Downloaded archives	82
A.3	MPEG-7 VCE dataset	83
A.3.1	Downloaded data	83
	Glossary	87
	Acronyms	89
	Bibliography	91
	Curriculum vitae	99