

Stefan Florczyk

Robot Vision

Video-based Indoor Exploration with
Autonomous and Mobile Robots



**WILEY-
VCH**

WILEY-VCH Verlag GmbH & Co. KGaA

Stefan Florczyk
Robot Vision

Stefan Florczyk

Robot Vision

Video-based Indoor Exploration with
Autonomous and Mobile Robots



**WILEY-
VCH**

WILEY-VCH Verlag GmbH & Co. KGaA

Author

Dr. Stefan Florczyk

Munich University of Technology
Institute for Computer Science
florczyk@in.tum.de

Cover Picture

They'll Be More Independent, Smarter and More
Responsive

© Siemens AG, Reference Number: SO CT 200204

■ All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.:

applied for

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Bibliographic information published by

Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at
<<http://dnb.ddb.de>>.

© 2005 WILEY-VCH Verlag GmbH & Co. KGaA,
Weinheim

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – nor transmitted or translated into machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Printed in the Federal Republic of Germany.

Printed on acid-free paper.

Typesetting Kühn & Weyh, Satz und Medien,
Freiburg

Printing betz-druck GmbH, Darmstadt

Bookbinding Litges & Dopf Buchbinderei GmbH,
Heppenheim

ISBN 3-527-40544-5

Dedicated to my parents

Contents

List of Figures *XI*

Symbols and Abbreviations *XV*

1	Introduction	1
2	Image Processing	9
2.1	Color Models	10
2.2	Filtering	11
2.2.1	Kalman Filter	11
2.2.2	Gabor Filter	13
2.2.3	Application of the Gabor Filter	16
2.3	Morphological Image Processing	22
2.3.1	The Structuring Element	22
2.3.2	Erosion	23
2.3.3	Dilation	23
2.4	Edge Detection	24
2.5	Skeleton Procedure	28
2.6	The Segmentation of Image Regions	28
2.7	Threshold	29
3	Navigation	33
3.1	Coordinate Systems	33
3.2	Representation Forms	36
3.2.1	Grid-based Maps	36
3.2.2	Graph-based Maps	37
3.3	Path Planning	38
3.3.1	Topological Path Planning	38
3.3.2	Behavior-based Path Execution	39
3.3.3	Global Path Planning	39
3.3.4	Local Path Planning	40
3.3.5	The Combination of Global and Local Path Planning	40

3.4	The Architecture of a Multilevel Map Representation	42
3.5	Self-localization	43
4	Vision Systems	47
4.1	The Human Visual Apparatus	47
4.1.1	The Functionality	47
4.1.2	The Visual Cortex	48
4.2	The Human Visual Apparatus as Model for Technical Vision Systems	49
4.2.1	Attention Control	50
4.2.2	Passive Vision	51
4.2.3	Active Vision	51
4.2.4	Space-variant Active Vision	52
4.3	Camera Types	53
4.3.1	Video Cameras	53
4.3.2	CCD Sensors	53
4.3.3	Analog Metric Cameras	55
5	CAD	57
5.1	Constructive Solid Geometry	57
5.2	Boundary-representation Schema (B-rep)	58
5.3	Approximate Models	59
5.3.1	Octrees	60
5.3.2	Extended Octrees	60
5.3.3	Voxel Model	61
5.4	Hybrid Models	62
5.5	Procedures to Convert the Models	62
5.6	The Use of CAD in Computer Vision	63
5.6.1	The Approximation of the Object Contour	64
5.6.2	Cluster Search in Transformation Space with Adaptive Subdivision	66
5.6.3	The Generation of a Pseudo-B-rep Representation from Sensor Data	71
5.7	Three-dimensional Reconstruction with Alternative Approaches	74
5.7.1	Partial Depth Reconstruction	74
5.7.2	Three-dimensional Reconstruction with Edge Gradients	75
5.7.3	Semantic Reconstruction	77
5.7.4	Mark-based Procedure	83
6	Stereo Vision	87
6.1	Stereo Geometry	87
6.2	The Projection of the Scene Point	90
6.3	The Relative Motion of the Camera	92
6.4	The Estimation of the Fundamental Matrix B	93
6.5	Image Rectification	95
6.6	Ego-motion Estimation	97
6.7	Three-dimensional Reconstruction by Known Internal Parameters	98

6.8	Three-dimensional Reconstruction by Unknown Internal and External Parameters	98
6.8.1	Three-dimensional Reconstruction with Two Uncalibrated Cameras	98
6.8.2	Three-dimensional Reconstruction with Three or More Cameras	100
6.9	Stereo Correspondence	105
6.9.1	Correlation-based Stereo Correspondence	106
6.9.2	Feature-based Stereo Correspondence	106
6.10	Image-sequence Analysis	109
6.11	Three-dimensional Reconstruction from Image Sequences with the Kalman Filter	110
7	Camera Calibration	113
7.1	The Calibration of One Camera from a Known Scene	114
7.1.1	Pinhole-camera Calibration	114
7.1.2	The Determination of the Lens Distortion	116
7.2	Calibration of Cameras in Robot-vision Systems	118
7.2.1	Calibration with Moving Object	120
7.2.2	Calibration with Moving Camera	121
8	Self-learning Algorithms	123
8.1	Semantic Maps	124
8.2	Classifiers for Self-organizing Neural Networks	125
9	OCR	129
10	Redundancy in Robot-vision Scenarios	133
10.1	Redundant Programs for Robot-vision Applications	134
10.2	The Program	135
10.2.1	Looking for a Rectangle	136
10.2.2	Room-number Recognition	137
10.2.3	Direct Recognition of Digits	138
10.2.4	The Final Decision	139
10.3	The Program Flow	140
10.4	Experiment	142
10.5	Conclusion	144
11	Algorithm Evaluation of Robot-vision Systems for Autonomous Robots	147
11.1	Algorithms for Indoor Exploration	148
11.1.1	Segmentation with a Gabor Filter	150
11.1.2	Segmentation with Highpass Filtering	152
11.1.3	Object Selection with a Band Filter	153
11.1.4	Object Detection with the Color Feature	153
11.1.5	Edge Detection with the Sobel Filter	155
11.2	Experiments	156
11.3	Conclusion	157

12	Calibration for Autonomous Video-based Robot Systems	159
12.1	Camera Calibration for Indoor Exploration	160
12.2	Simple Calibration with SICAST	160
12.2.1	Requirements	160
12.2.2	Program Architecture	161
12.3	Experiments	164
12.4	Conclusion	165
13	Redundant Robot-vision Program for CAD Modeling	167
13.1	New CAD Modeling Method for Robot-vision Applications	168
13.1.1	Functionality	168
13.1.2	Program Architecture	172
13.2	Experiment	182
13.3	Conclusion	184
	Bibliography	185
	Index	193

List of Figures

- Figure 1** The architecture of a video-based robot navigation software 3
- Figure 2** The one-dimensional Gabor filter [13] 14
- Figure 3** The variation of Gabor wavelength and spectrum factor [13] 15
- Figure 4** The wooden cube within a set of other objects [14] 17
- Figure 5** The regulator circle [14] 17
- Figure 6** The approximation of simple cells with a Gabor filter [16] 18
- Figure 7** Sequence of test images with two Gabor families [14] 19
- Figure 8** The wooden cube under different conditions [14] 20
- Figure 9** Gripping precision with the Gabor approach [14] 21
- Figure 10** Some structuring elements [18] 22
- Figure 11** The erosion of the set A 23
- Figure 12** The dilation of a set A 24
- Figure 13** Example for a convolution 25
- Figure 14** Edge detection with the Sobel operator 27
- Figure 15** The image `RawSegmentation` 30
- Figure 16** The six degrees of freedom [33] 34
- Figure 17** Conversion from locations [33] 35
- Figure 18** Coordinate systems for a mobile robot 35
- Figure 19** An allocentric map [38] 36
- Figure 20** A topological map [38] 38
- Figure 21** Sensorial situations of a robot [46] 41
- Figure 22** Example of a view graph with global and local edges [46] 42
- Figure 23** The architecture of a multilevel map representation [38] 43
- Figure 24** An example of the Monte Carlo localization [47] 44
- Figure 25** An abstract view of the human visual apparatus [49] 47
- Figure 26** Layers of the visual cortex 48
- Figure 27** Abstract presentation of a technical vision system [49] 49
- Figure 28** The pinhole camera model [33, 63] 54
- Figure 29** Model of a pinhole camera recording an aerial view [65] 56
- Figure 30** Representation of a three-dimensional model with CSG model [68] 57
- Figure 31** Representation of a three-dimensional object with B-rep [68] 59
- Figure 32** Three types in the octree [67] 60
- Figure 33** Additional types in extended octrees [72] 60

- Figure 34** Different voxel models [67] 61
- Figure 35** The $\Theta(lc)$ representation of a polygon contour [93] 66
- Figure 36** Cluster search in a two-dimensional transformation space [93] 67
- Figure 37** Algorithm for the cluster search [93] 68
- Figure 38** The binormals of an object contour [93] 71
- Figure 39** The preprocessing of a textured object [95] 76
- Figure 40** Stripe projection [95] 76
- Figure 41** Three-dimensional analysis with the projection of model data into image data [65] 78
- Figure 42** Semantic net for a building model [65] 79
- Figure 43** Matching between image data and model data [65] 79
- Figure 44** Segmentation of streets and areas in three steps [65] 80
- Figure 45** The filling up of regions with triangles [65] 81
- Figure 46** Building of a tetrahedron [33] 84
- Figure 47** Geometry in stereo vision [63] 88
- Figure 48** Canonical stereo configuration [99] 88
- Figure 49** Stereo geometry in canonical configuration [63] 89
- Figure 50** Epipoles e and e' in left and right image [63] 93
- Figure 51** Mismatches between corresponding points [63] 95
- Figure 52** Rectified images to support the matching process [102] 96
- Figure 53** Scene observed from three cameras [63] 101
- Figure 54** Plane with optical center F and scene point X [63] 102
- Figure 55** One trilinear relation [63] 103
- Figure 56** The calculation of the cyclopean separation [63] 107
- Figure 57** State-space representation [93] 112
- Figure 58** Coordinate systems of a robot-vision system [33] 113
- Figure 59** Relation between the coordinates of the projected point [63] 116
- Figure 60** Reference object with points in snake form [33] 119
- Figure 61** Six positions of the robot [33] 120
- Figure 62** Seven positions of the robot's camera [33] 121
- Figure 63** Semantic map [49] 124
- Figure 64** Classification with SOM [49] 125
- Figure 65** Connection between SOM and ACG [49] 127
- Figure 66** The modification of the threshold in the ACG [49] 128
- Figure 67** Ambiguity in character recognition [114] 129
- Figure 68** Characters that are stuck together 130
- Figure 69** Merging within a character 130
- Figure 70** Similar numerals 131
- Figure 71** A numeral that is not closed 131
- Figure 72** Direct recognition of a room number 139
- Figure 73** Program flow 141
- Figure 74** An image of poor quality 142
- Figure 75** An image with an acute angle to the doorplate 143
- Figure 76** A dark image 143
- Figure 77** A bright image 144

- Figure 78** Class design of an object segmentation algorithm 149
- Figure 79** Image I_O from a corridor 150
- Figure 80** Gabor filtered image I_G 151
- Figure 81** Highpass filtered image I_H 152
- Figure 82** Fire extinguisher in a threshold image 154
- Figure 83** The three-dimensional calibration object 161
- Figure 84** Program architecture 162
- Figure 85** Bookshelf that is represented with an ICADO model 170
- Figure 86** Class architecture of RICADO 173
- Figure 87** Report for an examined image 180
- Figure 88** Table that was imaged from different distances 182
- Figure 89** Performance measurements 183

Symbols and Abbreviations

A	Region
AO	Association operators
AW	Window
B	Blue channel
BC_j	Plane image contour j
BE	Base elements
C	CCD array
CE	Complex elements
CO	Contour
D	Projection matrix
E	Essential matrix
$E[x]$	Expected value for random vector x
$E(A)$	Erosion of a pixel set A
F	Focal point
$F\{f(x)\}$	Fourier transform for function $f(x)$
G	Green channel
$G(f_x)$	Fourier transform for Gabor filter
H	The horizontal direction vector of the Fz_{CHL} projection equation
$H(t)$	Observation matrix
I	Image
J	Jacobi matrix
K	Matrix
L_i	The Jacobi matrix of the observation equation at point in time i
L	Left-hand side
L	The vertical direction vector of the Fz_{CHL} projection equation
M	Rotation matrix
N	Unit matrix
O	Origin
P	Covariance matrix
P_k^+, P_{k+1}^-	The update and prediction of the covariance matrix, respectively
Q	Index set
Q_i	The covariance matrix of process noise at point in time i

R	Red channel
R	Right-hand side
S	Structuring element
SE	Set
T	Rotation
$Tf(x)$	Vectorial transformation function
U	Disparity
U	Color channel
V	Color channel
W	Skew symmetric matrix
$W(A)$	Function that calculates the width of a region A
X	Point in three-dimensional space
Y	Brightness
Z	Diagonal matrix
Z_i	Motion matrix at point in time i
a	Length
ao	Association operator
b	Focal length
c	Constant
d	Distance
d_i	The binormal vector of a polygon with index i
dr_{\max}	Maximal recursion depth by the cluster search
e_x^C, e_y^C, e_z^C	The column vectors of transformation matrix $H_{B_C, C}$ in camera coordinates
e	Epipole
f	Result
f_x	Frequency measured at X-axis
f_{x_m}	Middle frequency measured at X-axis
g	Epipolar line
g_x	The size of a pixel in X direction
g_y	The size of a pixel in Y direction
$g(x)$	One-dimensional Gabor filter
h_i	The number of bisections per axis i in the transformation space
h	Principal point offset
i	Imaginary number
ip	Vectorial input parameter
l	Location
lc	Contour length
lg	The local goodness of single element i in the transformation space
m	Node
n	Normal
nf	Norm factor
o	Octave
of	Offset
p	Pixel

$q(t)$	State vector
$\tilde{q}(t)$	Linear state valuer
$r(t)$	Random vector
r	Edge
s	Spectrum factor
se	Element in set SE
t	Time in general
u	Projection
v	Variable
$(x, y)_A$	Two-dimensional image affine coordinates
$(x, y)_I$	Two-dimensional image Euclidean coordinates
$(x, y)_S$	Two-dimensional sensor coordinates
$(x, y, z)_W$	Three-dimensional world coordinates
$(x, y, z)_{\bar{W}_{k+1}}$	The update of three-dimensional world coordinates at point in time $k + 1$
$(x, y, z)_{\bar{W}_k}^+$	The prediction of three-dimensional world coordinates at point in time k
A	The camera's line of sight
B	Fundamental matrix
$B(A)$	Function that calculates the area of a region A
X_A	Image affine coordinate system with axes X_A , Y_A , and Z_A
X_{BC_j}	Contour-centered coordinate system
X_C	Camera Euclidian coordinate system with axes X_C , Y_C , and Z_C
X_I	Image Euclidean coordinate system with axes X_I , Y_I , and Z_I
X_M	Robot coordinate system
X_S	Sensor coordinate system with axes X_S , Y_S , and Z_S
X_T	Transformation space with axes X_T , Y_T , and Z_T
X_W	World Euclidean coordinate system with axes X_W , Y_W , and Z_W
Δ	Difference
H	Homogeneous matrix
$H(A)$	Function that calculates the height of a region A
I	Matrix
K	Resolution power
$\Phi(t + 1; t)$	Transition matrix
$\Phi(t + 1; t)^*$	Estimated transition matrix
Γ	Disparity gradient
M	Matrix
N	Matrix
O	The vectorial output variable of a transformation
P	Segment
T	Threshold
Y	Ξ/K Quality
Ω	Tensor
Ξ	Field of vision
Ψ	Calibration matrix

Z	Matrix
α	Rotation applied to X-axis
β	Rotation applied to Y-axis
δ	Dilation
ε	Error value
ϕ	Activation profile
$\varphi(x)$	The local phase of a Gabor filter
γ	Rotation applied to Z-axis
η	Surface normal
ι	Intensity
κ	Quality factor
λ	Gabor wavelength
μ	Solution
μ	Learning rate
ν	Vector
$o(t)$	Observation vector
ϖ	Probability
θ	Angle of rotation
ϑ	Object size
ρ	Scale factor
σ	Standard deviation
τ	Translation
v	Neuron
ω	The element of a vector
$\omega(t)$	Random vector
ξ	Angle that is surrounded from two polygon areas
ψ	Distortion coefficient
ζ	Balance

1

Introduction

The video-based exploration of interiors with autonomous and mobile service robots is a task that requires much programming effort. Additionally, the programming tasks differ in the necessary modules. Commands, which control the technical basis equipment, must consider the reality of the robot. These commands activate the breaks and the actuation. The parts are basically included in the delivery. Often a mobile robot additionally possesses sonar, ultrasonic, and cameras, which constitute the perception function of the robot. The programming of such a mobile robot is a very difficult task if no control software comes with the robot. First, the programmer must develop the necessary drivers. As a rule the manufacturer includes a software library into the scope of the supply. This enables programs in a high-level language like C++ to be created very comfortably to control most or all parts of the robot's basic equipment. Typically operators are available. The user can transfer values to the arguments whose domain depends on the device that is to be controlled, and the admitted measurement. Operators, which enable rotary motions, may take values in degrees or radians. The velocity can be adjusted with values, which are specified in meters per second or yards per second. Video cameras are sometimes also part of a mobile robot's basic equipment, but further software and/or hardware must be acquired generally. A frame grabber is required. This board digitizes the analog signal of the camera. The gained digital image can then be processed with an image-processing library. Such a library provides operators for the image processing that can also be included into a high-level program. If the camera comes with the robot, the manufacturer provides two radio sets if the computer that controls the robot is not physically compounded with the robot. One radio set is necessary to control the robot's basic equipment from a static computer. The second radio set transmits the analog camera signals to the frame grabber. Nowadays, robots are often equipped with a computer. In this case radio sets are not necessary, because data transfer between a robot's equipment and a computer can be directly conducted by the use of cables. Additionally, a camera head can be used that connects a camera with a robot and enables software-steered panning and tilting of the camera. Mobile service robots use often a laser that is, as a rule, not part of a robot. They are relatively expensive, but sometimes the robot-control software provided involves drivers for commercial lasers.