

Lecture Notes in Mathematics

1948

Editors:

J.-M. Morel, Cachan

F. Takens, Groningen

B. Teissier, Paris

Frédéric Cao · José-Luis Lisani
Jean-Michel Morel · Pablo Musé
Frédéric Sur

A Theory of Shape Identification

Frédéric Cao
DxO Labs
3 rue Nationale
92100 Boulogne Billancourt, France
fcao@dxo.com

José-Luis Lisani
Dep. Matemàtiques i Informàtica
University Balearic Islands
ctra. Valldemossa km.7,5
07122 Palma de Mallorca, Balears
Spain
jose Luis.lisani@uib.es

Jean-Michel Morel
CMLA, Ecole Normale
Supérieure de Cachan
61 av. du Président Wilson
94235 Cachan Cédex, France
morel@cmla.ens-cachan.fr

Pablo Musé
Instituto de Ingeniería Eléctrica
Facultad de Ingeniería
Julio Herrera y Reissig 565
11300 Montevideo, Uruguay
pmuse@fing.edu.uy

Frédéric Sur
Loria Bat. C - projet Magrit
Campus Scientifique - BP 239
54506 Vandoeuvre-lès-Nancy Cédex
France
sur@loria.fr

ISBN 978-3-540-68480-0 e-ISBN 978-3-540-68481-7
DOI 10.1007/978-3-540-68481-7

Lecture Notes in Mathematics ISSN print edition: 0075-8434
ISSN electronic edition: 1617-9692

Library of Congress Control Number: 2008927359

Mathematics Subject Classification (2000): 62C05, 62G10, 62G32, 62H11, 62H15, 62H30, 62H35, 68T10, 68T45, 68U10, 91E30, 94A08, 94A13, 94B70

© 2008 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: WMX Design Bender

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

Recent years have seen dramatic progress in shape recognition algorithms applied to ever-growing image databases. They have been applied to image stitching, stereo vision, image retrieval, image mosaics, solid object recognition and video and web shape retrieval. More fundamentally, the ability of humans and animals to detect and recognize shapes is one of the enigmas of perception. Digital images and computer vision methods open new ways to address this enigma.

Given a dictionary of digitized shapes and a previously unobserved digital image, the aim of shape recognition algorithms is to know whether some of the shapes in the dictionary are present in the image. This book describes a complete method that starts from a query image and an image database and yields a list of the images in the database containing the query shapes.

Technically speaking there are two main issues. The first is extracting invariant shape descriptors from digital images. Indeed, a shape can be seen from various angles and distances and in various lights. A shape can even be partially occluded by other shapes and still be identifiable. Because the extraction step is so crucial, three acknowledged shape descriptors, SIFT (Scale-Invariant Feature Transform), MSER (Maximally Stable Extremal Regions) and LLD (Level Line Descriptor) will be introduced.¹

The second issue is deciding whether two shape descriptors are identifiable as the *same shape* or not. This decision process will derive from a unique paradigm, called the Helmholtz principle. For each decision a background model is introduced. Then one decides whether an event of interest (such as the presence of a shape in the image) has occurred if it has a very low probability of occurring by chance in the background model. Thus from the statistical viewpoint shape identification goes back to *multiple hypothesis testing*.

A shape descriptor is recognized if it is not likely to appear by chance in the background model. At a higher complexity level, a group of shape descriptors is recognized if its spatial arrangement could not occur just by chance. These two decisions

¹ In a recent review paper on affine invariant recognition written by a pool of experts, SIFT and MSER were actually acclaimed as the best shape descriptors [122].

rely on simple stochastic geometry and eventually compute a false alarm number for each shape descriptor. The lower this number, the more secure the identification. In that way most familiar simple shapes or images can be reliably identified. Many realistic experiments show false alarm rates ranging from 10^{-5} to less than 10^{-300} .

All in all these lecture notes prove that many shapes can indeed be identified. For these shapes one needs no *a priori* model and no training, just one sample of the shape and what statisticians call a *background model*, or a *null model*. In the case of shape recognition, the term background is to be taken to the letter. By the Helmholtz principle a shape is conspicuous if and only if it cannot be generated by the image background on which it is perceived. The background model is therefore easily learnt from the image database itself.

The above description should not be taken to suggest that the shape recognition problem is solved. The methods described only apply to solid shapes and not to deformable shapes. They only deal with individual shapes and images such as logos or paintings, and not with wide classes of objects such as all humans, all cats or all cars. This latter problem is known as *categorization* and is still widely open to research.

The authors are indebted to their collaborators for many important comments and corrections, particularly to Andrés Almansa, Yann Gousseau and Guoshen Yu. David Mumford and another anonymous referee made valuable comments which reshaped the book. All experiments were done using the public software MegaWave (authors: Jacques Froment and Lionel Moisan). The SIFT method is also public and downloadable.

The present theory was mainly developed at the Centre de Mathématiques et Leurs Applications, at ENS Cachan, at the Universitat de les Illes Balears and at IRISA, Rennes. It was partially financed for the past eight years by the Centre National d'Etudes Spatiales, the Centre National de la Recherche Scientifique, the Office of Naval research (grant N00014-97-1-0839) and the Ministère de la Recherche (project ISII-RNRT), and the Ministerio de Educación y Cultura (project MTM2005-08567). Special thanks to Bernard Rougé and Wen Masters for their great interest and support. We are indebted to Nick Chriss for numerous stylistic corrections.

Frédéric Cao
 José Luis Lisani
 Jean-Michel Morel
 Pablo Musé
 Frédéric Sur

Contents

1	Introduction	1
1.1	A Single Principle	1
1.2	Shape Invariants and Consequences	4
1.2.1	Shape Distortions	4
1.3	General Overview	9
1.3.1	Extraction of Shape Elements	9
1.3.2	Shape Element Encoding	11
1.3.3	Recognition of Shape Elements	11
1.3.4	Grouping	12
1.3.5	Algorithm Synopsis	12

Part I Extracting Image boundaries

2	Extracting Meaningful Curves from Images	15
2.1	The Level Lines Tree, or Topographic Map	15
2.2	Matas <i>et al.</i> Maximally Stable Extremal Regions (MSER)	17
2.3	Meaningful Boundaries	18
2.3.1	Contrasted Boundaries	18
2.3.2	Maximal Boundaries	19
2.4	A Mathematical Justification of Meaningful Contrasted Boundaries	21
2.4.1	Interpretation of the Number of False Alarms	21
2.5	Multiscale Meaningful Boundaries	26
2.6	Adapting Boundary Detection to Local Contrast	27
2.6.1	Local Contrast	29
2.6.2	Experiments on Locally Contrasted Boundaries	30
2.7	Bibliographic Notes	32
2.7.1	Edge Detection	32
2.7.2	Meaningful Boundaries vs. Haralick's Detector	33
2.7.3	Level Lines and Shapes	34
2.7.4	Tree of Shapes, FLST, and MSER	35
2.7.5	Extracting Shapes from Images	35

Part II Level Line Invariant Descriptors

3	Robust Shape Directions	41
3.1	Flat Parts of Level Lines	41
3.1.1	Flat Parts Detection Algorithm	42
3.1.2	Reduction to a Parameterless Method	43
3.1.3	The Algorithm	44
3.1.4	Some Properties of the Detected Flat Parts	44
3.2	Experiments	45
3.2.1	Experimental Validation of the Flat Part Algorithm	45
3.2.2	Flat Parts Correspond to Salient Features	47
3.3	Curve Smoothing and the Reduction of the Number of Bitangent Lines	49
3.4	Bibliographic Notes	54
3.4.1	Detecting Flat Parts in Curves	54
3.4.2	Scale-Space and Curve Smoothing	57
4	Invariant Level Line Encoding	61
4.1	Global Normalization and Encoding	61
4.1.1	Global Affine Normalization	61
4.1.2	Application to the MSER Normalization Method	64
4.1.3	Geometric Global Normalization Methods	65
4.2	Semi-Local Normalization and Encoding	66
4.2.1	Similarity Invariant Normalization and Encoding Algorithm	67
4.2.2	Affine Invariant Normalization and Encoding Algorithm	70
4.2.3	Typical Number of LLDs in Images	71
4.3	Bibliographic Notes	73
4.3.1	Geometric Invariance and Shape Recognition	73
4.3.2	Global Features and Global Normalization	74
4.3.3	Local and Semi-Local Features	75

Part III Recognizing Level Lines

5	A <i>Contrario</i> Decision: the LLD Method	81
5.1	A <i>Contrario</i> Models	81
5.1.1	Shape Model or Background Model?	81
5.1.2	Detection Terminology	83
5.2	The Background Model	85
5.2.1	Deriving Statistically Independent Features from Level Lines	87
5.3	Testing the Background Model	89
5.4	Bibliographic Notes	91
5.4.1	Shape Distances	91
5.4.2	A <i>Contrario</i> Methods	92

6	Meaningful Matches: Experiments on LLD and MSER	93
6.1	Semi-Local Meaningful Matches	93
6.1.1	A Toy Example	94
6.1.2	Perspective Distortion	98
6.1.3	A More Difficult Problem	101
6.1.4	Slightly Meaningful Matches between Unrelated Images	104
6.1.5	Camera Blur	105
6.2	Recognition Relative to Context	113
6.3	Testing A <i>Contrario</i> MSER (Global Normalization)	116
6.3.1	Global Affine Invariant Recognition. A Toy Example	116
6.3.2	Comparing Similarity and Affine Invariant Global Recognition Methods	116
6.3.3	Global Matches of Non-Locally Encoded LLDs	118

Part IV Grouping Shape Elements

7	Hierarchical Clustering and Validity Assessment	129
7.1	Clustering Analysis	129
7.2	A <i>Contrario</i> Cluster Validity	131
7.2.1	The Background Model	131
7.2.2	Meaningful Groups	132
7.3	Optimal Merging Criteria	136
7.3.1	Local Merging Criterion	136
7.4	Computational Issues	140
7.4.1	Choosing Test Regions	140
7.4.2	Indivisibility and Maximality	142
7.5	Experimental Validation: Object Grouping Based on Elementary Features	143
7.5.1	Segments	144
7.5.2	DNA Image	146
7.6	Bibliographic Notes	148
8	Grouping Spatially Coherent Meaningful Matches	151
8.1	Why Spatial Coherence Detection?	151
8.2	Describing Transformations	153
8.2.1	The Similarity Case	153
8.2.2	The Affine Transformation Case	154
8.3	Meaningful Transformation Clusters	155
8.3.1	Measuring Transformation Dissimilarity	155
8.3.2	Background Model: the Similarity Case	157
8.4	Experiments	158
8.5	Bibliographic Notes	161

9	Experimental Results	167
9.1	Visualizing the Results	167
9.2	Experiments	168
9.2.1	Multiple Occurrences of a Logo	168
9.2.2	Valbonne Church	173
9.2.3	Tramway	175
9.3	Occlusions	177
9.4	Stroboscopic Effect	179

Part V The SIFT Method

10	The SIFT Method	185
10.1	A Short Guide to SIFT Encoding	185
10.1.1	Scale-Space Extrema	186
10.1.2	Accurate Key Point Detection	187
10.1.3	Orientation Assignment	188
10.1.4	Local Image Descriptor	189
10.1.5	SIFT Descriptor Matching	189
10.2	Shape Element Stability versus SIFT Stability	190
10.2.1	An Experimental Protocol	190
10.2.2	Experiments	191
10.2.3	Some Conclusions Concerning Stability	195
10.3	SIFT Descriptors Matching versus LLD A <i>Contrario</i> Matching	196
10.3.1	Measuring Matching Performance	198
10.3.2	Experiments	201
10.4	Conclusion	207
10.5	Bibliographic Notes	207
10.5.1	Interest Points of an Image	207
10.5.2	Local Descriptors	207
10.5.3	Matching and Grouping	208
11	Securing SIFT with A <i>Contrario</i> Techniques	209
11.1	A <i>Contrario</i> Clustering of SIFT Matches	209
11.2	Using a Background Model for SIFT	210
11.3	Meaningful SIFT Matching	214
11.3.1	Normalization	214
11.3.2	Matching	215
11.3.3	Choosing Sample Points	218
11.4	The Detection Algorithm	219
11.4.1	Experiments: Securing SIFT Detections	220
11.5	Bibliographic Notes	224

A	Keynotes	225
A.1	Cluster Analysis Reader's Digest	225
A.1.1	Partitional Clustering Methods	225
A.1.2	Iterative Methods for Partitional Clustering	227
A.1.3	Hierarchical Clustering Methods	228
A.1.4	Cluster Validity Analysis and Stopping Rules	231
A.2	Three classical methods for object detection based on spatial coherence	235
A.2.1	The Generalized Hough Transform	235
A.2.2	Geometric Hashing	236
A.2.3	A RANSAC-based Approach	237
A.3	On the Negative Association of Multinomial Distributions	239
B	Algorithms	243
B.1	LLD Method Summary	243
B.2	Improved MSER Method Summary	244
B.3	Improved SIFT Method Summary	245
	References	247
	Index	255