Fortschritt-Berichte VDI

VDI

Reihe 10

Informatik/
Kommunikation

Nr. 855

Dipl.-Ing. Arne Ehlers,
Hannover

Object Detection using Feature Mining in a Distributed Machine Learning Framework

Institut für Informationsverarbeitung
www.tnt.uni-hannover.de

Object Detection using Feature Mining in a Distributed Machine Learning Framework

Von der Fakultät für Elektrotechnik und Informatik

der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

Doktor-Ingenieur

(abgekürzt: Dr.-Ing.)

genehmigte

Dissertation

von

Dipl.-Ing. Arne Ehlers

geboren am 24. November 1976 in Itzehoe.

2016

Referent:	Prof. Dr.-Ing. B. Rosenhahn
Korreferent:	Prof. Dr.-Ing. E. Reithmeier
Vorsitzender:	Prof. Dr.-Ing. J. Ostermann
Tag der Promotion:	09.11.2016

Fortschritt-Berichte VDI

Reihe 10

Informatik/
Kommunikation

Dipl.-Ing. Arne Ehlers,
Hannover

Nr. 855

Object Detection using Feature Mining in a Distributed Machine Learning Framework

Institut für Informationsverarbeitung
www.tnt.uni-hannover.de

Ehlers, Arne
Object Detection using Feature Mining in a Distributed Machine Learning Framework
Fortschr.-Ber. VDI Reihe 10 Nr. 855. Düsseldorf: VDI Verlag 2017.
162 Seiten, 72 Bilder, 10 Tabellen.
ISBN 978-3-18-385510-0, ISSN 0178-9627,
€ 62,00/VDI-Mitgliederpreis € 55,80.

Keywords: Object Detection – Feature Mining – Fractal Features – Data Augmentation – Machine Learning – Adaptive Boosting – Distributed Computing

This dissertation addresses the problem of visual object detection based on machine-learned classifiers. A distributed machine learning framework is developed to learn detectors for several object classes creating cascaded ensemble classifiers by the Adaptive Boosting algorithm. Methods are proposed that enhance several components of an object detection framework:
At first, the thesis deals with augmenting the training data in order to improve the performance of object detectors learned from sparse training sets.
Secondly, feature mining strategies are introduced to create feature sets that are customized to the object class to be detected. Furthermore, a novel class of fractal features is proposed that allows to represent a wide variety of shapes.
Thirdly, a method is introduced that models and combines internal confidences and uncertainties of the cascaded detector using Dempster's theory of evidence in order to increase the quality of the post-processing.

Bibliographische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter http://dnb.ddb.de abrufbar.

Bibliographic information published by the Deutsche Bibliothek
(German National Library)
The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at http://dnb.ddb.de.

Als Manuskript gedruckt. Printed in Germany.

ISSN 0178-9627

ISBN 978-3-18-385510-0

Acknowledgement

The present thesis originates from my work at the Institut für Informationsverarbeitung (TNT) of the Gottfried Wilhelm Leibniz Universität Hannover. Doing research at the institute has been a great experience. I have learned a lot about computer vision, machine learning, and object detection. But all this would not have been possible without the help of many people I would like to thank at this point.

I would like to express my deep gratitude to Prof. Dr.-Ing. Bodo Rosenhahn for giving me the opportunity to work in his group and for the supervision of my dissertation. His guidance, advises and feedback largely contributed to making this thesis a success.

I also like to thank Prof. Dr.-Ing. Eduard Reithmeier for being the second supervisor of my thesis and Prof. Dr.-Ing. Jörn Ostermann for being the chair of my defense committee.

Special thanks go to Florian Baumann. We made up a superb team by complementing each other perfectly in many tasks of research and development. I very much enjoyed our corporate work on many papers and participating several times together in conferences.

I would like to thank all colleagues at the institute for their help and support and their contribution to create such a brilliant atmosphere at the TNT. Especially, I would like to mention Matthias Reso, Stephan Preihs, Kai Cordes, Björn Scheuermann, Hanno Ackermann, and Marco Munderloh. Thank you very much for giving suggestions, criticisms, and encouragement in discussions and as well for the great time we spent together.

A very special thanks goes to my parents Hilde and Walter Ehlers who enabled me to study and always supported me.

This work has been partially funded by the DFG within the excellence cluster REBIRTH. The author gratefully acknowledges the support.

Contents

Abbreviations **VII**

Symbols and Notation **IX**

Abstract/Kurzfassung **XII**

1 Introduction **1**

2 Related Work and Data Sets **16**
2.1 Machine Learning for Visual Object Detection 17
2.1.1 Feature Provision 17
2.1.2 Learning Algorithms 19
2.2 Data Sets and Benchmarks 21

3 Fundamentals **28**
3.1 Common Features 29
3.1.1 Haar-like Features 29
3.1.2 Histograms of Oriented Gradients 31
3.1.3 From Features to Classifiers 33
3.2 Supervised Machine Learning 34
3.2.1 Adaptive Boosting 35
3.2.2 Viola and Jones Detection Framework 38
3.2.3 Margin Analysis 50
3.2.4 Variants of Boosting Algorithms 51
3.3 Data Analysis 52
3.3.1 Cluster Analysis 53
3.3.2 Principal Component Analysis 56
3.4 Detector Performance Measures 58

4 Distributed Machine Learning Framework **62**

5 Learning from Sparse Training Data **68**
5.1 Training Data Augmentation 70
5.2 Experimental Results 71
5.2.1 Experiments on Face Detection 71
5.2.2 Experiments on Cell Data Set 74

5.3 Discussion . . . 74

6 Fractal Integral Paths **76**
6.1 Boosted Fractal Integral Paths . . . 79
6.1.1 Fractals . . . 79
6.1.2 Fractal Features . . . 80
6.1.3 Fractal Properties . . . 80
6.1.4 Construction of Fractals . . . 82
6.1.5 Feature Types . . . 82
6.2 Experimental Results . . . 83
6.2.1 Face Detection . . . 83
6.2.2 Microscopic Cell Detection . . . 87
6.2.3 Training and Computing Time . . . 89
6.3 Discussion . . . 89

7 Multi-Feature Mining for Detector Learning **90**
7.1 2Rec Features . . . 93
7.2 Keypoint HOG Features . . . 95
7.3 Experimental Results . . . 99
7.3.1 Face Detection . . . 99
7.3.2 Lateral Car Detection . . . 110
7.3.3 Pedestrian Detection . . . 113
7.3.4 Insights into the Training Process . . . 118
7.4 Discussion . . . 120

8 Non-Maximum Suppression using Dempster's Theory of Evidence **121**
8.1 Merging Multiple Detections based on Dempster's Theory . . . 123
8.1.1 Cascaded Classifier . . . 123
8.1.2 Dempster-Shafer Theory of Evidence . . . 124
8.1.3 Joint Confidence based on Dempster-Shafer . . . 125
8.1.4 Confidence-based Detection Merging . . . 125
8.2 Experimental Results . . . 126
8.2.1 Face Detection . . . 127
8.2.2 Lateral Car Detection . . . 129
8.3 Discussion . . . 129

9 Conclusion **131**

A Lindenmayer Systems Defining Fractal Integral Paths **135**
A.1 L-System Defining Gosper Curve . . . 136
A.2 L-System Defining E-Curve . . . 136

Bibliography **138**

Abbreviations

1D	one dimensional
2D	two dimensional
3D	three dimensional
AdaBoost	Adaptive Boosting
ADAS	Advanced Driver Assistance Systems
AEB	Autonomous Emergency Braking
Bagging	Bootstrap Aggregating
CNN	Convolutional Neural Network
DS	Dempster-Shafer Theory of Evidence
Euro NCAP	European New Car Assessment Programme
$\mathbf{F_1}$	F-measure
FP	False Positive
FPGA	Field-Programmable Gate Array
FPPI	False Positives Per Image
FPPW	False Positives Per Window
FPR	False Positive Rate
GMM	Gaussian Mixture Model
HCI	Human-Computer Interaction
HMM	Hidden Markov Model
HOG	Histograms of Oriented Gradients
ICA	Independent Component Analysis
KPCA	Kernel Principal Component Analysis
KPHOG	Keypoint HOG
L-System	Lindenmayer System
MPI	Message Passing Interface
NMS	Non-Maximum Suppression

OpenCV	Open Source Computer Vision library
OpenMP	Open Multi-Processing
PCA	Principal Component Analysis
ROC	Receiver Operating Characteristic
SIFT	Scale-Invariant Feature Transform
SLIC	Simple Linear Iterative Clustering
SMD	Surface-Mount Device
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate

Symbols and Notation

General Symbols

p	probability
$\mathbf{x}$	training sample
y	label for a training sample
I	image matrix
μ_I	arithmetic mean (pixel) of an image I
N_{Pixel}	number of pixels in an image or detection window
u, v	image coordinates
X	random variable
$\mathrm{Var}(X)$	variance of random variable X
$\mathrm{E}(X)$	expected value of random variable X
i, j, n, r, s	control variables
N_{TP}	number of true positive classifications
N_{FP}	number of false positive classifications
N_{P}	number of positives in test set
N_{N}	number of negatives in test set

Symbols for Adaptive Boosting

$\mathcal{S}$	set of training examples
$N_{\mathcal{S}}$	number of training examples
$\mathcal{X}$	set of training samples
$\mathcal{Y}$	set of possible training labels
N_R	number of training rounds or weak classifiers
W	distribution over the training set $\mathcal{S}$
h	weak hypothesis or weak classifier
ϵ_r	training error in round r
β_r	mapping factor for round r
H	strong hypothesis
J	loss function

Symbols for Viola and Jones Learning Scheme

γ	feature
ρ	polarity of a weak classifier
θ	decision threshold of weak classifier
$N_{\mathcal{P}}$	number of positive training examples
$N_{\mathcal{N}}$	number of negative training examples
w	weight of training example
e_i	indicator function for classification of i-th training example
$\sum_{r=1}^{N_R} \alpha_r h_r(\mathbf{x})$	linear combination
α_r	weighting factor for weak classifier of round r
$\mathcal{P}$	set of positive training examples
$\mathcal{N}$	set of negative training examples
$w_{r,\mathcal{P}}$	sum of weights of positive training set $\mathcal{P}$ in round r
$w_{r,\mathcal{N}}$	sum of weights of positive training set $\mathcal{N}$ in round r
$\hat{w}_{r,\mathcal{P}}$	sum of positive example weights before in sorted list including the current example in round r
$\hat{w}_{r,\mathcal{N}}$	sum of negative example weights before in sorted list including the current example in round r

Symbols for Viola and Jones Cascade Scheme

t	minimum acceptable true positive rate or detection rate per stage
T	true positive rate or detection rate of a cascaded classifier
f	maximum acceptable false positive rate per stage
F	false positive rate of a cascaded classifier
F_{Target}	target false positive rate of a cascaded classifier
τ	decision threshold of strong classifier
N_S	number of stages in a cascaded classifier
N_{Rs}	number of weak classifiers in s-th cascade stage
h_{r_s}	weak classifier of round r in s-th cascade stage
H_s	strong classifier or ensemble of weak classifiers in s-th cascade stage
α_{rs}	weighting factor for r-th weak classifier in s-th cascade stage
τ_s	decision threshold of the strong classifier or ensemble of weak classifiers in s-th cascade stage

Symbols for Learning from Sparse Training Data

Ψ	mean of the training samples
$\overline{\mathbf{x}}$	difference vector of samples to mean training data Ψ

$\overline{X}$	zero-mean data matrix of shifted training samples $\overline{\mathbf{x}}$
C	covariance matrix
ϕ	non-linear transformation into a higher dimensional space
k	kernel function $k(\mathbf{x}_i,\mathbf{x}_j) = \langle\phi(\mathbf{x}_i),\phi(\mathbf{x}_j)\rangle$
λ	parameter controlling strength of projection into PCA-space
U	matrix of left-singular vectors
Σ	diagonal matrix of singular values
V	matrix of right-singular vectors
Λ	negative training example to be projected

Symbols for Fractal Integral Paths

$\mathcal{G} = (\mathcal{V},\omega,\mathcal{R})$	grammar defining a Lindenmayer system
$\mathcal{V}$	alphabet of a Lindenmayer system
ω	string of symbols from $\mathcal{V}$ defining the initial state
$\mathcal{R}$	set of production rules
X, Y, F, $+$, $-$	symbols used in alphabet $\mathcal{V}$

Symbols for Dempster-Shafer Theory of Evidence

$\emptyset$	empty set
Ω	frame of discernment or hypotheses set
$\wp(\Omega)$	power set
A	hypothesis, $A \in \wp(\Omega)$
m	mass function
$\otimes$	Dempster's rule of combination

Symbols for Non-Maximum Suppression using Evidence Theory

D_i	i-th candidate detection
ι, κ	scale in u and v-dimension
D'_j	refined merged detection for j-th cluster L_j of candidate detections
$\mathcal{L}$	partition of the set of candidate detections
L	subset or cluster of candidate detections
Γ_j	confidence of j-th detection cluster

Abstract

An important task in visual recognition systems, aiming on the extraction and interpretation of information in images or videos, is the detection of objects. In this process, all instances of a specified object class are requested to be localized in the visual input data. Object detection is essential for many applications that require a more comprehensive scene understanding, like advanced driver assistance systems or self-driving cars. The utilized object detectors are often created by machine learning algorithms that follow the paradigm of learning from examples. In a computational expensive training process, the algorithms learn the characteristics and visual appearance of the object class from training examples but the created detector has to work very fast and efficiently. Frequently, the object characteristics are not directly extracted from the observations but from a feature representation of the input data that gives a guiding principle on how to identify distinctive structures.

This thesis addresses the problem of visual object detection based on machine-learned classifiers. A distributed machine learning framework is developed to learn detectors for several object classes creating cascaded ensemble classifiers by the Adaptive Boosting algorithm. Methods are proposed that enhance several components of an object detection framework to improve its performance:

At first, the thesis deals with augmenting the training data in order to improve the performance of object detectors learned from sparse training sets. This problem frequently arises in industrial applications when highly specialized detectors are learned for e.g. quality assurance.

Secondly, methods are proposed to enhance the feature set that is utilized in the detector learning and its application. Feature mining strategies are introduced in order to create feature sets that are customized to the object class to be detected. By adapting to distinctive object structures, more representative features are assembled in a set of manageable size that enables an efficient detector learning. Furthermore, a novel class of fractal features is proposed that allows to represent a wide variety of shapes.

Thirdly, improvements are proposed to the post-processing that is performed after applying the learned detector to further work up its output. Commonly, this involves the assignment of confidences, merging detections that are very close to each other and dropping detections having low confidence. A method is introduced that models and combines internal confidences and uncertainties of the cascaded detector using Dempster's theory of evidence in order to increase the quality of the post-processing.

Keywords: Object Detection, Feature Mining, Fractal Features, Data Augmentation, Machine Learning, Adaptive Boosting, Distributed Computing

Kurzfassung

Die Objektdetektion ist eine wichtige Teilaufgabe im maschinellen Sehen, welches die Extraktion von Informationen aus Bildern oder Videos und deren Interpretation zum Ziel hat. Hierbei sollen sämtliche Instanzen einer Objektklasse in den visuellen Eingangsdaten lokalisiert werden. Die Detektion von Objekten ist eine elementare Voraussetzung für weitergehende Verfahren wie Fahrerassistenzsysteme oder selbstfahrende Autos, die eine umfassendere Wahrnehmung ihrer Umgebung erfordern. Die eingesetzten Objektdetektoren sind häufig durch maschinelle Lernalgorithmen erstellt worden, die dem Paradigma des Lernens anhand von Beispielen folgen. Der Algorithmus lernt hierbei in einem rechenintensiven Trainingsprozess das charakteristische Aussehen der Objektklasse anhand von Trainingsbeispielen. Der erstellte Detektor hingegen muss sehr schnell und effizient arbeiten. Häufig werden die Objektcharakteristiken nicht direkt aus den wahrgenommenen Eingangsdaten sondern aus einer Merkmalsdarstellung extrahiert, die Richtlinien zur Identifizierung markanter Strukturen vorgibt.

Diese Dissertation befasst sich mit der visuellen Objektdetektion durch maschinell gelernte Klassifikatoren. Ein verteiltes maschinelles Lernsystem ist entwickelt worden, um mit Hilfe des Adaptive Boosting Algorithmus Ensemble-Klassifikatoren für unterschiedliche Objektklassen anzulernen. Es werden Verfahren zur Verbesserung verschiedener Komponenten eines Objektdetektionssystems vorgestellt, um die Detektionsleistung des Gesamtsystems zu erhöhen:

Als Erstes beschäftigt sich diese Arbeit mit der Anreicherung der Trainingsdaten, um die Leistung von Detektoren zu steigern, welche auf kleinen Trainingsmengen angelernt werden. Diese Problematik tritt häufiger bei industriellen Anwendungen auf, wenn hoch spezialisierte Detektoren beispielsweise für die Qualitätssicherung erstellt werden sollen.

Der zweite Beitrag der Dissertation stellt Verfahren zur Verbesserung der Merkmalsmengen vor, die beim Anlernen eines Detektors und während der Detektion genutzt werden. Es werden Methoden zur gezielten Generierung von Merkmalsmengen entwickelt. Hierdurch können die Merkmalsmengen an die Charakteristiken der zu detektierenden Objektklasse angepasst werden, sodass eine Menge von aussagekräftigeren Merkmalen entsteht, die gleichzeitig überschaubar ist und somit ein effizientes Anlernen erlaubt. Weiterhin wird eine neue Klasse von Fraktalmerkmalen vorgestellt, die vielfältige Strukturen repräsentieren kann.

Drittens werden Verbesserungen für die Detektionsnachverarbeitung entwickelt. Üblicherweise werden den Detektionen in diesem Schritt Konfidenzen zugewiesen, nah beieinander gelegene Detektionen verschmolzen und Detektionen mit niedriger Konfidenz verworfen. Ein Verfahren wird vorgestellt, dass interne Konfidenzen und Unsicherheiten der Detektorkaskade mit Hilfe der Evidenztheorie modelliert und kombiniert, um die Qualität der Nachverarbeitung zu erhöhen.

Stichworte: Objektdetektion, Merksmalsextraktion, Fraktalmerkmale, Datenaugmentation, Maschinelles Lernen, Adaptive Boosting, Verteiltes Rechnen