Fahrzeugtechnik

Julian Schwehr

Gaze Target Tracking for Driver Assistance Systems

SHAKER
VERLAG

Gaze Target Tracking for Driver Assistance Systems

Dem Fachbereich
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung des akademischen Grades
eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von

Julian Jürgen Schwehr, M. Sc.

geboren am 27. August 1990 in Frankfurt am Main

Referent:	Prof. Dr.-Ing. J. Adamy
Korreferent:	Prof. Dr. rer. nat. H. Winner
Tag der Einreichung:	28. April 2020
Tag der mündlichen Prüfung:	15. Juli 2020

D17
Darmstadt 2020

Berichte aus der Fahrzeugtechnik

Julian Schwehr

Gaze Target Tracking for Driver Assistance Systems

D 17 (Diss. TU Darmstadt)

Shaker Verlag
Düren 2020

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.: Darmstadt, Techn. Univ., Diss., 2020

Printed in Germany.

ISBN 978-3-8440-7702-5
ISSN 0945-0742

Shaker Verlag GmbH • Am Langen Graben 15a • 52353 Düren
Phone: 0049/2421/99011-0 • Telefax: 0049/2421/99011-9
Internet: www.shaker.de • e-mail: info@shaker.de

Preface

This dissertation is the result of my work at the Control Methods and Robotics Lab of the Institute of Automatic Control and Mechatronics, TU Darmstadt. It was embedded in the PRORETA 4 project – a research cooperation with Continental. One of the first things I learned at the beginning of the research project was the origin of its name. The "proreta" was the officer at the bow on ancient Roman ships whose task was to keep an eye out for obstacles and shallows. The past five years have been an extraordinary journey on the sea called my life.

First of all, I want to thank my doctorate supervisor Prof. Jürgen Adamy for not only giving me the chance to go on this voyage but also for providing a pleasant working climate in his lab with the great possibility to sail wherever the wind of one's own ideas takes you. I am also grateful to Prof. Hermann Winner for his acceptance to act as second referee and even more for motivating and constructive feedback in the many PRORETA meetings providing fresh wind in one's sails.

I owe a special thanks to Volker Willert for all his time, support, ideas and confidence which helped me to sail around obstacles and shallows and not to capsize halfway on the trip.

Stefan Luthardt, Hien Dang, Maren Henzel and Nils Magiera were the best crew members I could wish for. The outstanding motivation to tackle any obstacle in the course of the project together made the daily work a real pleasure!

I am also deeply indebted to all Continental colleagues who contributed to the success of PRORETA and also this dissertation, namely Maximilian Höpfl, Benedikt Lattke, Christoph Wannemacher, Saman Khodaverdian, Ronald Bayer, Herbert Deckenbach, Johannes Eck, Alfred Eckert, Knut Ehm, Moritz Groh, Alexander Klotz, Ralph Lauxmann, Guido Mayer-Arendt, Karsten Michels, Rex Schilasky, Christian Thur, Manfred Wilck and all other colleagues who supported the project in the background.

To all my past colleagues at the Control Methods and Robotics Lab I want to say thanks for the great time and discussions giving fresh impetus to my work every day. Thanks to Hanno Winter and Moritz Bühler for proof reading this manuscript at an early stage and providing valuable

feedback.

I am very grateful to my parents who have been and still are supporting me in every step in the journey of my life no matter where I'm heading to. Thanks for being stable anchors in my life.

Last but not least I would like to give a special thanks to Katharina who started this project as my girlfriend, became my wife and ended the journey as mother of our son. Thanks for the everlasting support, for always cheering me up at times where nothing seemed to work and reminding me that there are more important things in life than this thesis.

Tettnang, in September 2020 Julian Schwehr

Car il suffit pour y voir clair de changer de perspective.

Antoine de Saint-Exupéry

Contents

Abbreviations and Symbols

Abbreviations

ABS	Anti-lock Braking System
Acc	Accuracy
ACC	Adaptive Cruise Control
AD	Automated Driving
ADAS	Advanced Driver Assistance System
ADF	Assumed Density Filter
DBN	Dynamic Bayesian Network
EEG	Electroencephalogram
EKF	Extended Kalman Filter
ESC	Electronic Stability Control
ETG	Eye-Tracking Glasses
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GPB	Generalized Pseudo-Bayesian
GUI	Graphical User Interface
HET	Head-Eye-Tracking
HMI	Human-Machine Interface
HMM	Hidden Markov Model
IMM	Interacting Multiple Model
IR	Infrared
IS	Intersection
KF	Kalman Filter
KL	Kullback-Leibler
LDS	Linear Dynamic System
LGS	Linear Gaussian System
LiLo	Looking in and Looking out
LKA	Lane Keeping Assist
MH	Multiple Hypothesis
MHMM	Multi-Hypothesis Multi-Model
MHMMP	Multi-Hypothesis Multi-Model tracking with gaze motivated parameter selection
MHMMPS	Multi-Hypothesis Multi-Model tracking with gaze motivated parameter selection and reduced sampling

ML	Maximum Likelihood
MLE	Maximum Likelihood Estimation
MM	Multiple Model
MOT	Multiple Object Tracking
NCC	Normalized Cross Correlation
pdf	probability density function
PERCLOS	Percentage of eye closure
PF	Particle Filter
POI	Point of Interest
PoR	Point of Regard
Pr	Precision
Re	Recall
ROC	Receiver Operating Characteristic
TH	Threshold
TN	True Negative
TP	True Positive
TPR	True Postive Rate
TTC	Time To Collision
TTI	Time To Intervention
UKF	Unscented Kalman Filter
VO	Visual Odometry
VRU	Vulnerable Road User

Notation

x, X	Scalar
$\boldsymbol{x}$	Column vector
$\boldsymbol{x}^\top$	Row vector
$\boldsymbol{x}^{0:k}$	Sequence of vectors $\boldsymbol{x}^{0:k} = \{\boldsymbol{x}^0, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^k\}$
$\boldsymbol{x}^{0:t}$	Time sequence of vectors $\boldsymbol{x}^{0:t} = \{\boldsymbol{x}^0, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^t\}$
$\boldsymbol{x}^k$	kth vector of vector sequence $\boldsymbol{x}^{0:k}$
$\boldsymbol{x}^t$	Vector at time t of time sequence of vectors $\boldsymbol{x}^{0:t}$
$\{\boldsymbol{x}, \boldsymbol{y}\}^{0:t}$	short notation for $(\boldsymbol{x}^{0:t}, \boldsymbol{y}^{0:t})$
$\boldsymbol{X}$	Matrix
$\boldsymbol{X}^\top$	Transpose of matrix $\boldsymbol{X}$
$\boldsymbol{X}^{-1}$	Inverse of matrix $\boldsymbol{X}$
$\mathbf{0}$	Zero vector or matrix
$\mathbf{1}$	Matrix of ones

$\bar{\boldsymbol{x}}$	3D homogeneous coordinates of 2D point $\boldsymbol{x}$ (only in the context of projective geometry)
$\bar{\boldsymbol{X}}$	4D homogeneous coordinates of 3D point $\boldsymbol{X}$ (only in the context of projective geometry)
$f(\cdot)$	Scalar function
$\boldsymbol{f}(\cdot)$	Vector function
$\int_{\boldsymbol{x}}(\cdot)d\boldsymbol{x}$	Integration over the whole range of $\boldsymbol{x}$, e. g. if $\boldsymbol{x} \in \mathbb{R}^n$, then $\int_{\boldsymbol{x}}(\cdot)d\boldsymbol{x} = \int_{x_1} \dots \int_{x_n}(\cdot)dx_1 \dots dx_n$
$\int_{\boldsymbol{x}^{0:k}}(\cdot)d\boldsymbol{x}^{0:k}$	Integration over the sequence of vectors $\boldsymbol{x}^{0:k}$, i. e. $\int_{\boldsymbol{x}^{0:k}}(\cdot)d\boldsymbol{x}^{0:k} = \int_{\boldsymbol{x}^0} \dots \int_{\boldsymbol{x}^k}(\cdot)d\boldsymbol{x}^0 \dots d\boldsymbol{x}^k$
$p(\boldsymbol{x})$	Probability density function if $\boldsymbol{x}$ is a continuous random vector or probability mass function if $\boldsymbol{x}$ is a discrete random vector
$p(\boldsymbol{x}, \boldsymbol{y})$	Joint pdf of $\boldsymbol{x}$ and $\boldsymbol{y}$
$p(\boldsymbol{x}\|\boldsymbol{y})$	Conditional pdf of $\boldsymbol{x}$ given $\boldsymbol{y}$
$p(x_j)$	probability that a discrete binary random vector $\boldsymbol{x}$ is in state j, i. e. $p(x_j) = p(x_j = 1)$
$q(\boldsymbol{x}\|\boldsymbol{\theta})$	Approximating distribution with parameters $\boldsymbol{\theta}$
$\mathcal{N}$	Normal distribution

Important Functions and Transformations

$p(\boldsymbol{x}^k\|\boldsymbol{z}^{1:k})$	Filtering distribution or current belief of $\boldsymbol{x}^k$ given $\boldsymbol{z}^{1:k}$
$p(\boldsymbol{x}^k\|\boldsymbol{x}^{k-1})$	Transition density
$p(\boldsymbol{z}^k\|\boldsymbol{x}^k), \ell(\boldsymbol{z}^k\|\boldsymbol{x}^k)$	Measurement likelihood
$\mathcal{N}(x\|\mu, \sigma)$	Normal pdf of a random variable x with mean μ and variance σ
$\mathcal{N}(\boldsymbol{x}\|\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Multivariate normal pdf of a random vector $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\mathrm{KL}(p(\boldsymbol{x})\|\|q(\boldsymbol{x}))$	Kullback-Leibler divergence between distributions $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$
$\mathrm{E}[\boldsymbol{x}]$	Expectation of random vector $\boldsymbol{x}$

Symbols

Latin Capital Letters

$\boldsymbol{A}$	System matrix of discrete-time linear dynamic system
$\boldsymbol{B}$	Input matrix of discrete-time linear dynamic system
$\boldsymbol{C}$	Measurement matrix of discrete-time linear dynamic system
K	number of samples
K'	number of free space samples
$\boldsymbol{K}$	Kalman gain matrix; Camera calibration matrix
$\boldsymbol{P}$	State error covariance matrix
$\boldsymbol{Q}$	Process noise covariance matrix
$\boldsymbol{R}$	Measurement noise covariance matrix
$\boldsymbol{R}_{cv}$	Rotation matrix between vehicle coordinate system and camera coordinate system
$\boldsymbol{X}_{g,c}$	3D point in camera coordinate system
$\boldsymbol{X}_{g,v}$	3D point in vehicle coordinate system
$\boldsymbol{X}_{\mathrm{PoR}}$	3D Point of Regard
Z_d	Depth of the scene at image point $\boldsymbol{x}_g$ obtained from stereo disparity map
Z_g	Depth of $\boldsymbol{X}_{g,c}$ in camera coordinates

Latin Lowercase Letters

b	Posterior belief
c	Normalization constant
d	Distance between point $\boldsymbol{x}$ and gaze origin $\boldsymbol{x}_o$ in 2D vehicle coordinates
d_{fs}	Interpolation distance between free space sample points
$\boldsymbol{f}$	Static environment measurement variable
g	Gaze yaw angle measurement variable
i	Running index, used at time $t-1$
j	Running index, used at time t
k	Running index, used for time sequence and sample points
l	Running index, used for particle filter sample points
$\boldsymbol{m}$	Dynamic objects measurement variable

n	State vector dimension; number of dynamic objects in object list
p	Input vector dimension
p_{ij}	Transition probability from model i to model j
q	Measurement vector dimension
r	Number of prediction models
$\boldsymbol{r}$	B-Spline curve with curve parameter s
s	Curve parameter of B-spline
$s^\star$	Curve parameter of B-spline of relevant spline points
$\boldsymbol{s}$	Discrete binary switching variable
$\boldsymbol{t}_{cv}$	Translation vector between vehicle coordinate system and camera coordinate system
$\boldsymbol{u}$	Input vector
$\boldsymbol{v}$	Measurement noise vector
$\boldsymbol{v}_{\text{rel}}$	Relative velocity to ego-vehicle
$\boldsymbol{w}$	Process noise vector
$\boldsymbol{x}$	State vector; spatial coordinate
$\boldsymbol{x}_{\text{a}}$	State vector of area of attention
$\boldsymbol{x}_{\text{fs}}$	Free space sample point
$\boldsymbol{x}_g$	2D image coordinates (reprojection) of 3D point $\boldsymbol{X}_{g,c}$
$\boldsymbol{x}_o$	gaze origin, i. e. location of the driver's eyes
$\boldsymbol{x}_{\text{P}}$	B-spline control point
$\boldsymbol{x}_{\text{PoR}}$	Point of Regard in 2D image coordinates
z_c	Weighting factor of normal distribution obtained from multiplication of two normal distributions
$\boldsymbol{z}$	Measurement vector

Greek Letters

α	Gaze yaw angle (heading)
α_{ij}	Temporal switching weight when switching from target i to target j
β	Gaze pitch angle (heading)
β_{ij}	Spatial switching weight when switching from target i to target j
γ_{ij}	Process noise vector when switching from target i to target j
$\boldsymbol{\Gamma}_{ij}$	Process noise covariance matrix of γ_{ij}

δ_0	Constant model parameter in sigmoid function of spatial switching weights
ϵ	Weighting of points which do not belong to free space sample points
η	Normalization constant
θ	Angular difference between measured gaze yaw direction and 2D point in vehicle coordinates
θ_{s_i}	Angular difference between measured gaze yaw direction and closest point of object i
$\boldsymbol{\theta}$	Parameters of approximating distribution $q(\boldsymbol{x}\|\boldsymbol{\theta})$
$\boldsymbol{\theta}^{\star}$	Parameters of best approximation $q(\boldsymbol{x}\|\boldsymbol{\theta})$ of $p(\boldsymbol{x})$ in terms of the KL-Divergence
κ	Constant model parameter in sigmoid functions of switching weights
λ	Sample weight; mean sample weight
Λ	Mode likelihood function
$\boldsymbol{\mu}$	Mean vector
$\boldsymbol{\mu}_{oj}$	Measured position of object j
ν	Measurement noise variable of gaze measurement
π_i	Mode probability of model i
π_{ij}	Merging probability
$\mathbf{\Pi}$	Model transition probability matrix
$\mathbf{\Pi}_0$	Canonical projection matrix
ρ	Weighting of free space sample points
σ	Variance of gaze likelihood
$\mathbf{\Sigma}$	Covariance matrix
$\mathbf{\Sigma}_{oj}$	Covariance matrix with Eigenvalues and Eigenvectors such that major axes of the 90 % covariance ellipse correspond to width, length, and heading of object j
τ_0	Constant model parameter in sigmoid function of temporal switching weights
φ	Measured heading angle of objects in object list

Abstract

Despite many supporting systems, so-called advanced driver assistance systems (ADAS), human error is still by far the main cause of traffic accidents. In the development of new driver assistance concepts, systems and functions monitoring the driver while driving and classifying their behavior in the driving context are therefore increasingly coming to the fore. In this context, this dissertation deals with the question what the driver perceived in their environment. For this purpose, the information of the environment model has to be merged with measured gaze data. Given a precise calibration of the individual sensors, visual fixations of the driver on road users are modeled.

Based on the realization that simple geometric approaches cannot answer this question of visual fixations precisely enough, characteristics of human gaze behavior are identified and integrated as model knowledge into a probabilistic tracking approach. This tracking model considers every object which is classified as a dynamic object and thus as a potential road user by the vehicle's environment perception module as a possible hypothesis for the driver's current visual attention target. In addition, two different motion models of eye movements for fixations and saccades are integrated, so that the estimation of the gaze target can follow the special dynamics of human gaze and recognize specific connected time spans. The advantage of this novel resulting Multi-Hypothesis Multi-Model (MHMM) filter is the confidence which is characteristic to probabilistic approaches, indicating the probability of each object being fixated by the driver.

A challenge is the evaluation of such new algorithms. For the statement which object the driver actually visually fixates, ground truth information is necessary. However, this cannot be covered by questionnaires. For this reason, a reference data set is created in which the recordings of the remote eye-tracking system installed in the vehicle are extended with the data of wearable eye-tracking glasses. With the help of these recordings, different model approaches are now compared on a quantitative and not only qualitative basis.

The prototypical City Assistant System, which was co-developed as part of this work, shows how the newly gained information about the driver's gaze behavior can be incorporated into new assistance concepts. It adapts its warning and recommendation cascade in urban intersection scenarios to the driver's driving style and gaze behavior. Through this orientation towards the driver's need for support, the City Assistant System contributes to higher acceptance of warning and recommending systems and ultimately to increased road safety.

Kurzfassung

Trotz vieler unterstützender Systeme, sogenannter Fahrerassistenzsysteme (FAS), sind menschliche Fehler immer noch mit großem Abstand die Hauptursache für Verkehrsunfälle. Bei der Entwicklung von neuen Fahrerassistenzkonzepten rücken daher verstärkt Systeme und Funktionen in den Vordergrund, die den Fahrer während der Fahrt beobachten und sein Verhalten im Fahrkontext einordenen und bewerten. In diesem Rahmen behandelt die vorliegende Dissertation die Frage, was der Fahrer in seinem Umfeld wahrgenommen hat. Hierzu sind die Informationen des Umfeldmodells mit den Messdaten der Blickrichtung zu fusionieren. Eine präzise Kalibrierung der einzelnen Sensoren vorausgesetzt werden visuelle Fixationen des Fahrers auf Verkehrsteilnehmern modelliert.

Basierend auf der Erkenntnis dass einfache geometrische Ansätze diese Frage nach visuellen Fixationen nicht klar genug beantworten können, werden zunächst Eigenschaften des menschlichen Blickverhaltens identifiziert und als Modellwissen in einen probabilistischen Trackingansatz integriert. Dieses Trackingmodell berücksichtigt jedes Objekt, welches von der Umfelderfassungssensorik des Fahrzeugs als dynamisches Objekt und damit als potentieller Verkehrsteilnehmer engestuft wird, als mögliche Hypothese für das aktuelle visuelle Aufmerksamkeitsziel des Fahrers. Zusätzlich sind für Fixationen und Sakkaden der Augenbewegungen zwei verschiedene Bewegungsmodelle integ-riert, sodass die Schätzung des Aufmerksamkeitsziels der speziellen Dynamik des menschlichen Blicks folgen und gezielt zusammenhängende Zeitspannen erkennen kann. Der Vorteil dieses neuen resultierenden Multi-Hypothesen Multi-Modell (MHMM) Filters besteht in der für probabilistische Ansätze charakteristischen Konfidenz, die für jedes Objekt angibt, gerade vom Fahrer angesehen zu werden.

Eine Herausforderung besteht in der Bewertung solch neuer Algorithmen. Für die Aussage, welches Objekt der Fahrer tatsächlich visuell fixiert, sind Referenzwerte notwendig, die nicht über Fragebögen abgedeckt werden können. Aus diesem Grund wird ein Referenzdatensatz erstellt, bei dem die Aufnahmen des entfernten, im Fahrzeug verbauten Eye-Tracking-Systems mit den Daten einer tragbaren Eye-Tracking-Brille erweitert werden. Mit Hilfe dieser Aufnahmen werden verschiedene Modellansätze nun quantitativ

und nicht mehr nur qualitativ miteinander verglichen.

Wie die neu gewonnene Information über das Blickverhalten des Fahrers in neue Assistenzkonzepte einfließen kann, zeigt der prototypische City Assistant, welcher im Rahmen dieser Arbeit mitentwickelt wurde. Dieser passt seine Warn- und Empfehlungskaskade in innerstädtischen Kreuzungsszenarien an den Fahrstil und das Blickverhalten des Fahrers an. Durch diese Orientierung am Unterstützungsbedarf des Fahrers leistet der City Assistant einen Beitrag zu höherer Akzeptanz von warnenden und empfehlenden Systemen und letztlich zu höherer Verkehrssicherheit.