Autonomes Fahren: Eine kritische Beurteilung der technischen Realisierbarkeit

Tobias Haist
Universität Stuttgart, Institut für Technische Optik, Pfaffenwaldring 9, 70569 Stuttgart, Germany

Abstrakt: In dieser kurzen Übersicht werde ich darstellen, warum in absehbarer Zukunft vollautonomes Fahren in ausreichender Qualität nicht technisch realisierbar sein wird. Entscheidend ist dabei, dass für eine nicht vernachlässigbare Anzahl an Situationen ein vergleichsweise gutes Szenenverständnis notwendig ist und wir beim Stand der Technik keine Idee haben, wie dieses Szenenverständnis realisierbar ist.

Datum: 09.09.2016

1 Einführung

Vollautonome Fahrzeuge versprechen eine Reihe von deutlichen Vorteilen gegenüber den heute menschlichen Fahrzeugführern. Aufgrund der schnellen Reaktionszeit und dem Vermeiden menschlicher Fehler erhofft man sich eine Verbesserung der Verkehrssicherheit bzw. eine Verringerung schwerwiegender oder gar tödlicher Unfälle. Auch der Kraftstoffverbrauch (und damit der Schadstoffausstoß) lässt sich durch eine optimale Fahrweise reduzieren. Ältere Verkehrsteilnehmer könnten durch autonome Fahrzeuge weiterhin uneingeschränkt mobil bleiben und natürlich kann jeder Fahrer von der Entlastung durch das autonome Fahrzeug, insbesondere bei langen Fahrzeiten und/oder bei Nacht profitieren. Sobald erst die Mehrzahl der Fahrzeuge autonom fahren ergeben sich weitere Vorteile, denn eine Vernetzung der Fahrzeuge kann die Unfallwahrscheinlichkeit natürlich massiv senken.

Die Motivation für das autonome Fahren ist also klar und es ist scheinbar nur eine Frage der Zeit, wann das autonome Fahren beim Endkunden ankommen wird. Für die deutsche Automobilindustrie scheint es dementsprechend unabdingbar zu sein, hier führend bei der Entwicklung (und nachfolgenden Markteinführung) zu partizipieren. Aus meiner Sicht eine (unter Umständen teure) Fehleinschätzung.

In diesem Artikel werde ich — entgegen der scheinbar vorherrschenden Expertenmeinung — begründen, warum das vollautonome Fahrzeug weder in fünf noch in zehn oder zwanzig Jahren großflächige Realität auf unseren Straßen werden wird. Dabei werde ich mich nicht auf die rechtlichen und ethischen Gründe, die dem autonomen Fahren entgegen stehen, stützen, sondern die Fragestellung aus rein technischer Sicht beleuchten. Meine Sicht ist dabei eine Sicht von außen. Ich arbeite nicht an autonomen Fahrzeugen und nicht im Bereich des Bildverstehens.

Die Argumentationslinie, die im folgenden detailliert wird, kann in aller Kürze folgendermaßen zusammengefasst werden: Für einen geringen Anteil kritischer Fahrsituationen ist ein komplexes Bild- bzw. Szenenverstehen notwendig. Dieses Verständnis wird mit bekannten Methodiken der Bildverarbeitung nicht erzielbar sein. Die Aufgabenstellung ist für nichtlernende Verfahren zu komplex, für lernende Verfahren ist der Lernvorgang aufgrund der nahezu unbegrenzten Situationsvielfalt nicht beherrschbar (Beispieldaten und Rechenzeit).

Einschränkend muss gesagt werden, dass es im folgenden explizit um *vollautonomes* Fahren im *heutigen* und beliebigen Straßenverkehr geht. Andere, zukünftige Verkehrskonzepte, z.B. Fahren in der Kolonne mit einem (menschlichen) Hauptfahrer oder auch komplett vernetzte Fahrsituationen, führen zu einer erheblichen Vereinfachung für das maschinelle Sehen.

In diesem Beitrag geht es alleinig um die Leistung, die aktuell menschlichen Fahrern abverlangt wird und eben die Frage, ob diese Leistung technisch in absehbarer Zukunft erzielbar sein wird. Explizit sei ebenfalls betont, dass die hier dargestellte Skepsis sich *nicht generell* auf maschinelles Sehen im Verkehrswesen, z.B. im Sinne von Fahrerassistenzsystemen, bezieht.

Abschnitt 2 stellt zunächst dar, was die Grundaufgabe des autonomen Fahrens ist. Eine kurze Darstellung der einsetzbaren Sensorik (Abschnitt 3) zeigt, dass — auch wenn hier noch Defizite bestehen — doch prinzipiell die Sensorik nicht das Problem darstellt bzw. in diversen Situationen dem menschlichen Fahrer sogar deutlich überlegen sein kann.

Abschnitt 4 verdeutlicht dann, dass gegenüber einem rein reaktiven Vorgehen und vergleichsweise einfachen Entscheidungsprozessen teilweise ein doch komplexes Szenenverständnis notwendig ist. Hierfür sind verschiedene Dinge notwendig, eine Basisvoraussetzung ist allerdings eine leistungsfähige Objekterkennung. Die Abschnitte 4, 5 und 6 beschreiben Ansätze und den Stand der Technik, vor allem aber auch die prinzipiellen Probleme, die hierbei auftreten. Im Abschnitt 7 werden weitere Probleme des vollautonomen Fahrens, die zunächst unabhängig von der Bildverarbeitung sind, kurz angesprochen.

Die von mir vorgebrachte Kritik ist natürlich nicht grundsätzlich neu und so sind natürlich auch typische Argumentationslinien, die üblicherweise gegen die (scheinbar unseriöse) Kritik vorgebracht werden, bekannt. Auf einige typische Argumente werde ich in Abschnitt 8 eingehen bevor ich in Abschnitt 9 Gründe für die scheinbar vorherrschende Meinung zum autonomen Fahren darlege. Abschnitt 10 behandelt die Frage, ob eine weitere Forschung vor dem Hintergrund der Probleme sinnvoll ist und schließlich werde ich im Abschnitt 11 mit einem kurzen Ausblick schließen.

2 Was ist notwendig?

Jedes autonom handelnde Objekt in unserer Umwelt, sei es ein Mensch, eine Katze oder auch ein autonomes Fahrzeug, muss für ein sinnvolles Verhalten fortlaufend folgende Fragen beantworten:

1. Welches bzw. was sind die relevanten Objekte in der Szene? (Identifikation)
2. Wo sind diese Objekte im dreidimensionalen Raum (absolut und relativ zu mir)? (Detektion)
3. Wohin werden die Objekte sich bewegen? (Vorhersage)
4. Welches Verhalten von mir ist in der Situation sinnvoll? (Handlungsanweisung)

Notwendig ist hierfür die Lösung verschiedener Grundaufgaben. Die Umgebung muss mittels Sensorik und Zusatzinformation (Kartographie[1], GPS, Datenbanken) erfasst werden. Die relevanten Objekte der Szene, ihre Lage im dreidimensionalen Raum werden erkannt. Eine Einschätzung des zu erwartenden Verhaltens der Objekte wird durchgeführt und schließlich wird eine Entscheidung hinsichtlich des eigenen weiteren Verhaltens getroffen.

[1] Hochauflösende Karten sind bei aktuellen Projekten zum autonomen Fahren besonders wichtig. [1, 2]. Die genaue Ermittlung der Eigenposition in Bezug auf die hochgenaue Karte stellt sicher, dass das Fahrzeug auf der Fahrbahn bleibt. Der Mensch geht hier natürlich anders vor und benötigt kein hochauflösendes Kartenmaterial, um das Fahrzeug zu lenken.

3 Sensorik

Von den Befürwortern des vollautonomen Fahrens wird (zu Recht) darauf verwiesen, dass die Maschine dem Mensch bei der sensorischen Erfassung deutlich überlegen sein *kann*. Das menschliche Auge hat beeindruckende Leistungsdaten: ca. 125 Millionen „Pixel", komprimiert bereits auf Netzhaut auf ca. 1 Million, sehr großes Blickfeld, beugungsbegrenzte Auflösung im Zentralbereich, extrem hohe Dynamik und extrem hohe Sensitivität, sehr gute Farbdifferenzierung[2], lange Lebensdauer, geringer Energieverbrauch, Redundanz durch zwei Augen, sehr schneller und stabiler Autofokus, sehr schnelle Kopf- und Augenbewegung[3] mit Aufmerksamkeitssteuerung sowie Auflösungsverbesserung durch Mikrosakkaden, geringes Gewicht und Größe (7.5 g, 5.5 cm^3), 80 Jahre wartungsfreie Lebensdauer, geringer Energieverbrauch, Stoßunempfindlichkeit.

Dennoch kann die Kombination moderner Sensoren das Auge in verschiedenen Bereichen (z.B. hochgenaue Abstandsbestimmung über große Entfernung) deutlich übertreffen. Eine Vielzahl von Sensoren bzw. Messgeräten kann heute helfen, eine Szene zu erfassen. Aktuell kommen vor allem folgende Sensorprinzipien beim autonomen Fahren zum Einsatz.

Konventionelle Kameras: Laterale Auflösungen liegen — je nach Objektfeld — im Bereich des menschlichen Sehens (1 Winkelminute) oder besser; mehrere Kameras sind kombinierbar. Die Bildaufnahme kann über die Blende (analog der Iris beim menschlichen Auge) oder aber die Belichtungszeit[4] an die Helligkeit der Szene angepasst werden. Nichtlineare Kameras erlauben — analog zum menschlichen Sehen — auch größere Dynamiken. Teure Sensoren erzielen dabei auch bei schlechten Lichtverhältnissen ein vergleichsweise hohes Signal-Rauschverhältniss (vorteilhaft für nachfolgende Bildverarbeitung). Schließlich ist hervorzuheben, dass eine Bildaufnahme im nahen Infrarotbereich oder gar SWIR (short wave infrared) bei Nebel oder auch Dunkelheit unter Umständen eine verbesserte Fernsicht bzw. Detektion warmer Objekte (Menschen) erlaubt [3]. Die Möglichkeiten sind beträchtlich, allerdings werden beim Stand der Technik in Fahrzeugen aus Kostengründen Kameras eingesetzt, die dem menschlichen Auge (s.o.) noch weit unterlegen sind.

Stereokameras: Letztlich können zwei oder mehr Kameras analog dem menschlichen Sehen eingesetzt werden, um a) Redundanz zu gewinnen, b) Dynamik oder Spektralbereiche zu erweitern, c) den Erfassungsbereich zu erweitern oder d) mittels Stereoalgorithmik auf die dreidimensionale Geometrie der Szene rückzuschließen. Die Genauigkeit der dreidimensionalen Bestimmung nimmt linear mit dem Abstand der Kameras voneinander (Stereobasis) zu. Für kurze und mittlere Entfernungen werden dabei beim Stand der Technik ausreichend gute Auflösungen erreicht. Allerdings muss hierfür die Kalibrierung sehr gut sein und vor allem stabil (starke Temperaturschwankungen im Automobil) über lange Zeit gegeben bleiben. Die Hauptschwierigkeit ist allerdings, bei einer hohen lateralen Auflösung und realen Szenen nur

[2] bis zu 1 nm bei monochromatischen Farben im gelb-grünen Spektralbereich

[3] Dies ist ein gewaltiger Vorteil, denn damit wird ein Rundumblick mit enormer Auflösung ermöglicht. Ein typisches Beispiel wo dies notwendig ist, ist das Warten in vorderster Reihe an Ampeln. Die Ampel ist dann teilweise nur unter einem extremen Blickwinkel sichtbar und natürlich muss dieser Blickwinkel durch die Kamerasensorik erfasst werden.

[4] Hierbei ist bei gepulsten Verkehrsschildern unter Umständen mit (technisch lösbaren) Problemen zu rechnen (flicker mitigation).

wenige Fehlstellen zu erzielen. Tippetts et al. haben viele verschiedene Algorithmen untersucht und kommen auf in der Größenordnung 10% Fehlpixel für die sogenannten Middlebury Testbilder. Aufwändigere Algorithmen können hier auch durchaus 4% erreichen (siehe [4] für einen Überblick oder [5] für ein gutes Einzelverfahren).

Hauptvorteil gegenüber anderen 3D Verfahren ist die rein passive Arbeitsweise, d.h. das Umgebungslicht wird genutzt. Eine aktive Ausleuchtung einer ausgedehnten Szenerie (z.B. im Entfernungsbereich von einigen zehn Metern) erfordert bei den notwendigen kurzen Belichtungszeiten vergleichsweise starke Lichtquellen, die potentiell problematisch hinsichtlich der Augensicherheit und teuer sind. Entsprechende Unterstützung durch aktive Beleuchtung ist aber denkbar. Die Hauptschwierigkeit liegt in der Verarbeitung der Stereosignale. Der Rückschluss auf die dreidimensionale Geometrie ist alles andere als trivial und weit entfernt von dem was der Mensch leistet [4, 5, 6].

Der Grund für die extreme Diskrepanz von technischem zu menschlichen Stereosehen liegt im Bild- bzw. Szenenverstehen (vgl. Abschnitt 4). Letztlich kann der Mensch auch bei reiner Betrachtung eines Einzelbilds sehr gut und fehlerfrei die räumlichen Positionen aller Objekte in der Szene angeben. Bei Einzelbildern sind hierfür die Perspektive in Kombination mit der Objekterkennung ausschlaggebend. Die Bildgröße eines Objekts nimmt linear mit der Entfernung ab, d.h. wenn ein Objekt bekannter Größe korrekt detektiert ist, dann kann aufgrund der Bildgröße auf die Entfernung rückgerechnet werden. Eine isolierte Betrachtung auf Pixelbasis oder kleinen Bildbereichen für die bei der traditionellen Stereovision eine Korrespondenzanalyse und daraus abgeleitet mit der Kalibrierung dann eine Entfernungsbestimmung durchgeführt wird, ist ungleich schwieriger. Der Mensch nutzt stattdessen vor allem diese exakte Objekterkennung und setzt dabei unterstützend im Bereich bis ca. 10 m die Stereodisparität ein. Eine deutliche Verbesserung der Robustheit technischer Stereosysteme kann auf zwei Wegen erreicht werden. Zum einen durch erweiterte Sensorik, also bei Stereo entweder Mehrkamerasysteme oder plenoptische Kameras oder aber durch ein verbessertes Szenenverstehen bzw. Objekterkennung. Erstere Variante erscheint beim Stand der Technik einfacher.

Ultraschall: Ultraschallsensoren, wie sie z.B. großflächig für Einparkhilfen zum Einsatz kommen, sind sehr gut geeignet, das Vorhandensein naher Objekte zu detektieren., Basis hierfür ist die Laufzeitmessung des Ultraschallsignals. Hiermit lassen sich Entfernungen selbst über mehrere Meter mit sub-Millimeter Genauigkeit erfassen. Die laterale Auflösung ist dabei aufgrund der großen Wellenlängen von Schall beugungsbedingt allerdings sehr schlecht [7]. Eine Identifikation von Objekten ist daher nicht möglich.

Radar: Laufzeitmessungen lassen sich natürlich nicht nur mittels Schall sondern auch mit elektromagnetischer Strahlung durchführen. Radarsensoren nutzen dabei Frequenzen von einigen 10 Gigahertz und erzielen damit bei Entfernungen von bis zu mehreren hundert Metern immerhin laterale Auflösungen im Bereich einiger zehn Zentimeter bis Meter. Je größer die Entfernung, desto größer muss auch das Objekt sein, um eine ausreichende Rückreflexion zu gewährleisten.

Für den Continental ARS 408-21 Premium Sensor wird beispielsweise eine Auflösung des Seitenwinkels von 1.6° angegeben [8]. Bei einem Abstand von 10 m entspricht dies einer

lateralen Auflösung von 10 m $\cdot \tan 1.6^\circ = 30$ cm. Eine genaue Identifikation eines Objekts (z.B. Fußgänger oder gleich großes Pappschild) ist also nicht möglich und kleine, schlecht reflektierende Objekte werden unter Umständen gar nicht detektiert.

Über den Dopplereffekt kann nicht nur die Entfernung sondern auch die Geschwindigkeit des Objekts (in Richtung des eigenen Fahrzeugs) sehr genau bestimmt werden.

Lidar: Die Laufzeitmessung elektromagnetischer Strahlung bei hohen Frequenzen (z.B. sichtbares Licht oder Licht im nahen Infrarot) kann potentiell deutlich höhere Auflösungen erzielen (letztlich sind theoretisch dieselben lateralen Auflösungen wie bei Kameras erzielbar, axiale Auflösungen im Millimeterbereich bei großen Entfernungen sind möglich und in jedem Fall deutlich besser als notwendig). Lidarsysteme scannen dabei die Szenerie und messen jeweils für den gescannten Strahl die Entfernung. Die effektive Auflösung ist daher vom Scanbereich in Kombination mit der Messdauer abhängig. Diese hängt wieder von der Lichtmenge, der Entfernung, der Reflektanz der Objekte und der Empfindlichkeit der Sensorik und der gewünschten Messunsicherheit ab. Die Reflektanz der Szene kann — wie im Sinne eines konventionellen Kamerabilds — ebenfalls ermittelt werden [1].

Oftmals werden parallelisierte Systeme mit vielen Strahlen eingesetzt, die Kosten für Lidarsysteme sind allerdings hoch und potentiell können Probleme mit der Augensicherheit auftreten.

Andere Lichtlaufzeitverfahren: Deutlich einfacher als konventionelle Lidarsysteme lassen sich auch andere, weniger genaue Lichtlaufzeitkameras realisieren [9, 10]. Letztlich ist durch die notwendige aktive Beleuchtung der Szene aber die maximale Entfernung deutlich begrenzt.

Aktive Verfahren, also Verfahren, die die Szene aktiv bestrahlen, können generell zu Problemen führen, sobald viele Fahrzeuge entsprechende Sensorik verwenden. Ein entgegenkommendes mit z.B. Radar ausgestattetes Fahrzeug kann durchaus zu einer Beeinträchtigung der eigenen Radarsensorik führen. Zwar können spezielle Radarcodes dafür sorgen, dass die Signale inkohärent zueinander bleiben, mindestens das Signal-Rausch-Verhältnis kann aber massiv verschlechtert werden.

Für eine Identifikation von Objekten sind in jedem Fall die optischen Verfahren wesentlich. Nicht-optische Verfahren liefern auf die Distanz keine ausreichende laterale Auflösung zur Objektidentifikation.

Technische Sensoren, die mit hoher Geschwindigkeit die Szene erfassen werden beim Stand der Technik — zumindest bei begrenzten Sensorkosten — noch durch eine Vielzahl relativ elementarer Störungen eingeschränkt.

Allerdings ist das kein prinzipielles Problem. Verbesserungen bei der Sensorik sind vergleichsweise direkt erzielbar und die Vorteile der potentiell deutlich höheren Geschwindigkeit und der höheren Genauigkeiten (z.B. Entfernungsbestimmung bei Lidar oder Sicht durch Nebel bei SWIR Kameras) lassen hoffen, dass in der Tat die Sensorik des autonomen Fahrzeugs der Sensorik des Menschen in naher Zukunft überlegen sein wird (bzw. eben in Teilbereichen bereits ist). Hier liegt eine klare Chance vor allem für Assistenzsysteme aber potentiell eben auch für autonome Fahrzeuge.

Die Leistung, die der Mensch beim Sehen (im Sinne des Gesamtvorgangs von sensorischer Erfassung und Verarbeitung) erzielt, ist allerdings phänomenal. Maßgeblich ist dabei aber vor allem die Verarbeitung, insbesondere das Verstehen der Szene. Hierdurch wird bei der Erfassung der Realität eine hohe Robustheit erreicht, die den technischen Systeme bisher fehlt.

4 Szenenverstehen

Die Vorteile der technischen Sensorik können bei einfachen Situationen klar und überzeugend ausgespielt werden. Nehmen Sie den Fall, dass ein Kind unerwartet vor ein Auto läuft. Die Reaktionsgeschwindigkeit des technischen Systems kann hier deutlich besser als die des Menschen sein. Die Bremsung beginnt früher (und wird dabei auch noch korrekt durchgeführt). Eine entsprechende Ausweichlenkbewegung erfolgt ohne nennenswerte Verzögerung.

In einer Vielzahl von Fällen kann so das autonome Fahrzeug in der Tat dem Mensch überlegen sein. Nämlich immer dann, wenn die primäre Schwierigkeit in der schnellen Reaktion auf einfache Reize bzw. Situationen liegt.

Allerdings reicht eine leichte Abwandlung der Situation, um für das Fahrzeug problematisch zu werden. Nehmen wir an, ein Reh läuft vor das Fahrzeug. Natürlich macht das einen Unterschied, denn eine Vollbremsung oder (falls der Bremsweg nicht ausreicht) ein riskantes Ausweichmannöver für ein Kind ist unumgänglich, während für das Reh unter Umständen die Risikoabwägung anders aussieht. Es ist also notwendig, dass die Objekterkennung z.B. ein Reh sicher von einem Kind unterscheidet. Aber auch ein brauner Hund muss von einem auf einem braunen Bobbycar fahrenden Kleinkind (im braunen Strampelanzug) unterschieden werden (s. Bild 1). Ein Fußgänger von einer Reklametafel etc.

Abbildung 1 Bobbycar und Hund. In welche Richtung soll “ausgewichen” werden, wenn eine Vollbremsung nicht mehr möglich ist?

Wesentlich für die Entscheidungsfindung wird nun plötzlich nicht mehr die Sensorgeschwindigkeit, sondern die Prozessierung der Sensorsignale wird aufwändig und bekommt einen erheblichen Einfluss.

Nehmen wir an, das Fahrzeug detektiert, dass der Bremsweg zur Vermeidung eines plötzlichen Hindernis nicht mehr ausreichend ist. Das Fahrzeug muss nun entscheiden, ob und wie es die Straße verlässt oder ein anderes Hindernis bevorzugt trifft oder eben das primäre Hindernis

umfährt. Hierzu ist es unerlässlich, die relevanten Objekte der Szene korrekt zu klassifizieren und die Szene insgesamt zu *„verstehen”*.

„Verstehen” kann dabei gleichgesetzt werden mit der Fähigkeit, alle *relevanten* Fragen in Bezug auf die Szene beantworten zu können. (Das ist natürlich eine schwammige Definition, aber eine exaktere Definition bringt an dieser Stelle keinerlei Vorteil.)

Ich will dies an einem einfachen und beliebigen Bild verdeutlichen. Betrachten Sie Abb. 2. Sie erfassen sofort und ohne Probleme die Szene. Was ist z.B. das rot umkreiste Objekt und wo ist es. Natürlich ein (parkendes) Auto und es ist vergleichsweise weit weg (sie könnten auch die Lage relativ zu anderen Objekten in der Szene ohne Probleme angeben). Und was ist das dahinter ? Sie sehen nur wenige Pixel, trotzdem ist klar, dass es ein weiteres Auto ist und auch die Lage im Raum und die Größe sind für Sie klar.

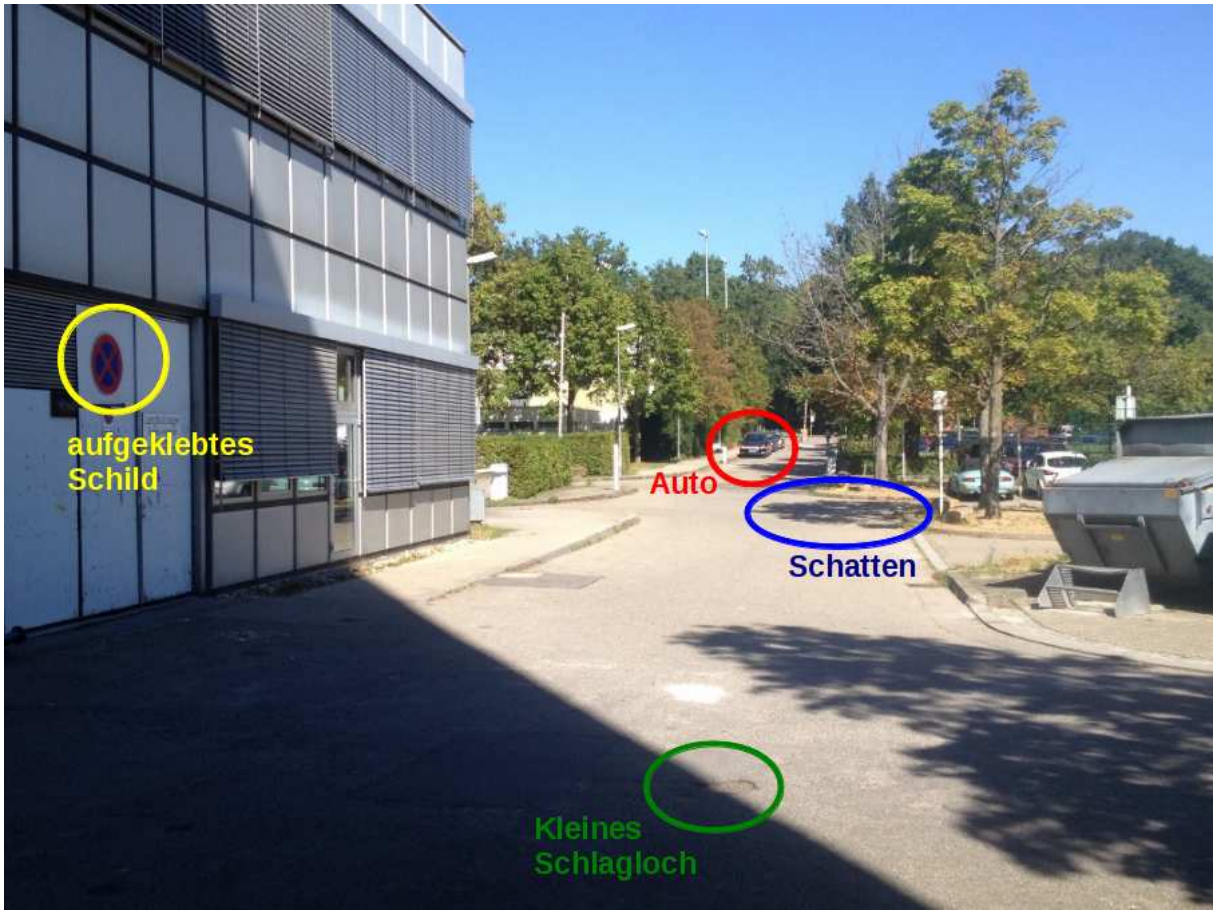


Abbildung 2 Eine beliebige Szene. Sie können jedes Objekt der Szene klar identifizieren und dreidimensional lokalisieren.

Sie sehen sofort, dass der grün umrandete Bereich ein Schlagloch in der Straße ist, Sie sehen, dass links ein Halteverbotsschild aufgeklebt und Sie sehen, dass es sich im Zentralbereich nur um Schatten aufgrund der Bäume auf der Straße handelt. Mit anderen Worten: Sie können jedes Objekt der Szenerie sofort und ohne Probleme direkt erkennen und im Raum anordnen. Sie könnten die Szenerie korrekt aus Pappmache nachbasteln. Sie verstehen alles Relevante (und noch viel mehr) der Szene. Das ist Szenenverstehen. Und wir haben keine Ahnung, wie wir diese Leistung technisch in dieser Form vollbringen können.

Abbildung 3 Bildverstehen am Beispiel einer weiteren Szene: Sie können z.B. klar den Tisch lokalisieren. Warum aber ist der Tisch als Tisch im zweidimensionalen Bild erkennbar? Letztlich erkennen Sie den Tisch nicht isoliert sondern als Bestandteil der gesamten Szene. Eine Einzeldetektion unabhängig vom Rest der Szene ist schwierig. Der Tisch wird in seinem Kontext detektiert.

Betrachten Sie ein zweites Bild (Abb. 3). Auch hier können Sie natürlich sofort die gesamte Szene verstehen. Lenken Sie Ihre Aufmerksamkeit auf den Tisch. Warum sehen Sie sofort, dass der Tisch ein Tisch ist? Was macht das „Tischsein" aus? Wie definieren Sie Tisch und vor allem: Wie würden Sie einem Computer beibringen, dass es sich um einen Tisch handelt? Falls Sie selbst schon programmiert haben, dann können sie vermutlich nachvollziehen, wie unlösbar diese Aufgabe für einen Programmierer ist. Tische können nämlich ganz unterschiedlich aussehen.

Wie erkennt ein Mensch den Tisch und warum ist es für den Mensch so einfach, den Tisch zu sehen egal in welcher Orientierung und Größe, verdeckt und unverdeckt? Wesentlich ist zunächst, dass der Mensch den Tisch nicht isoliert erkennt, sondern als Bestandteil der Gesamtszene. Der Tisch wird in seinem *Kontext* gesehen. Wenn ich dem Mensch nur die vom Tisch sichtbaren Bildpunkte präsentiere, dann wird eine Erkennung scheitern. Es kommt vielmehr darauf an, dass wir *alle* bzw. viele Objekte der Szene erkennen und zusätzliches *Wissen* über typische Szenen haben. So wird uns klar, dass die Objekte, die auf dem Tisch stehen eben Objekte sind, die auf einem Tisch stehen. Erst dadurch, dass wir die anderen Objekte erkennen, können wir erkennen, dass der Tisch ein Tisch ist. Umgekehrt erkennen wir die Objekte auf dem Tisch unter anderem dadurch, dass wir erkannt haben, dass es sich bei dem Tisch um einen Tisch handelt. Die Objekterkennung ist also kein isoliertes, lokales Problem sondern ein Problem, das die gesamte Szene betrifft. Dies ist eine wesentliche Erkenntnis.

Betrachten wir im folgenden das wichtige Problem der Detektion von Fußgängern (laut [11] sind in den USA bei 1/7 aller tödlichen Unfälle im Straßenverkehr Fußgänger involviert). Natürlich können Fußgänger ganz unterschiedlich aussehen (Kleidung, Größe, Ansicht, Position, Haartracht, Entfernung etc.) und sich unterschiedlich bewegen (rennend, stehend, mit Rollator laufend etc.). Die Fußgänger können teilweise durch andere Objekte (Ampeln, parkende Autos, andere Fußgänger etc.) verdeckt sein (siehe z.B. [12]). Vielleicht sitzen oder liegen sie sogar auf der Fahrbahn. Die Variabilität ist immens und die rein isolierte Betrachtung der Fußgängerbildpunkte reicht nicht aus, um eine robuste Entscheidung „Fußgänger/kein-Fußgänger" zu treffen.

Um es salopp zu sagen: Es reicht also ganz und gar nicht, nur Fußgänger zu detektieren, wenn man Fußgänger detektieren will. Wenn man Fußgänger von anderen Objekten, z.B. Werbetafeln, unterscheiden können will dann reicht es nicht einmal, Fußgänger und Werbetafeln erkennen zu können. Nein. In aller Regel wird eine robuste Erkennung von Fußgängern nur dann möglich sein, wenn die Szene in der näheren Umgebung des Fußgängers insgesamt korrekt erfasst wird.

Der Mensch schafft diese Unterscheidung mit hoher Qualität. Darüber hinaus wird er die Klasse Fußgänger weiter und ohne Probleme unterteilen können. Wir sehen z.B. ob es sich um einen Polizisten oder ein Kind oder einen Betrunkenen handelt. Auch das Verhalten können wir mit hoher Güte abschätzen, z.B: Wie groß ist die Gefahr, dass das Kind, gleich auf die Straße rennen wird?

4.1 Fußgängerdetektion mit dem Computer

Betrachten wir nun aber den Stand der Technik zu diesem Thema. Nur weil uns keine Idee kommt, wie wir denn eine robuste Klassifikation in Fußgänger und Nicht-Fußgänger dem Computer beibringen können, muss es ja nicht heißen, dass andere — vielleicht geniale — Programmierer das nicht könnten. Wir betrachten also hier exemplarisch, wie gut dieses wichtige Teilproblem des Bildverstehens von den Spezialisten auf diesem Gebiet bisher beherrscht wird.

Natürlich können Sie einwenden, dass die Programmierer bei Google, Daimler und Co. in Wirklichkeit viel weiter sind und die veröffentlichten Resultate nicht den wirklichen Stand der Technik repräsentieren. Das mag sein und wir können deshalb vielleicht — wohlwollend — einen zusätzlichen Bonus einräumen. (Andererseits sollten wir als Gesellschaft fordern, dass die Leistungsfähigkeit der Fahrzeuge, die autonom auf unseren Straßen rollen sollen, auch realistisch, nachprüfbar und klar definiert veröffentlicht wird, denn wie soll die Gesellschaft (oder der Gesetzgeber) sonst entscheiden, ob die Risiken für die Öffentlichkeit tragbar sind?)

Die Beurteilung der Leistungsfähigkeit für diese Teilaufgabe ist ein ziemlich komplexes Unterfangen und das macht die Diskussion darüber natürlich nicht einfacher. Man benötigt riesige Datenbanken von Bildern, die — letztlich vom Mensch — beurteilt wurden. Eine der beeindruckendsten Datenbanken für ziemlich beliebige Szenen (nicht für Fahrerassistenz) ist aktuell Microsofts COCO („Common Objects in Context") [13]. Sie besteht aus 328.000 Bildern mit insgesamt 2.500.000 Millionen gekennzeichneten Objekten aus 91 Objektkategorien (z.B. „Tisch"). Für die meisten der Kategorien sind damit mehrere tausend Testobjekte vorhanden, so dass die Datenbank zum Lernen aber auch zum Test der Leistungsfähigkeit verwendet werden kann. Viele weitere Bilddatenbanken für das Szenenverständnis wurden angelegt [13]. Im Vergleich zu den bekannteren (weil älteren) ImageNet und PASCAL Datenbanken sind in COCO mehr Bilder vorhanden, in denen die zu findenden Objekte in kompletten Szenen — also nicht isoliert — eingebunden sind. Das

macht das Erkennen einerseits schwieriger, zum anderen steht aber auch mehr *kontextuelle Information* zur Verfügung (z.B. *Gläser*, die auf einem *Tisch* in einem *Wohnzimmer* stehen). Für den Mensch ist das — wie bereits besprochen — wesentlich für die Gesamterkennung der Szene. Die Wahrscheinlichkeit, dass ein Objekt als Tisch erkannt wird steigt, wenn sich das Objekt im Wohnzimmer befindet und sich Gläser darauf befinden. Das Erkennen profitiert also von der Gesamtszene, vom Kontext. Allerdings eben nur, wenn der Kontext bei der Auswertung überhaupt ausreichend genutzt wird.

Soweit die Testdaten. Wie sollen wir aber nun die Leistung des Systems beurteilen? Hier sind verschiedene Metriken in Gebrauch. Machen Sie sich zunächst klar, dass die Erkennungsrate davon abhängt, wieviele falsch-positive Erkennungen zugelassen werden. Betrachten wir die beiden Extreme: Ich kann ohne Probleme ein Programm schreiben, dass garantiert alle Bilder mit Fußgängern findet. Dazu sage ich — ohne dass ich irgendetwas im Bild analysiere — einfach immer: „Da ist auf jeden Fall ein Fußgänger." Ich werde mit dieser Vorgehensweise keinen Fußgänger übersehen. Natürlich ist das wenig sinnvoll, denn gleichzeitig werde ich natürlich laufend scheinbare Fußgänger detektieren obwohl dort gar keine sind (falsch-positive Detektionen).

Auch der umgekehrte Fall ist direkt einsichtig: Ich kann ebenfalls ohne Probleme ein Programm schreiben, das es vermeidet, falschen Fußgängeralarm zu geben. Hierzu gibt mein Programm einfach immer aus, dass kein Fußgänger vorhanden ist. Ich werde also garantiert niemals fälschlich einen Fußgänger detektieren. Aber natürlich ist auch dieses Programm nutzlos, denn ich werde auch niemals einen echten Fußgänger finden.

Für ein sinnvolles System müssen wir also das System so einstellen, dass es ausgewogen (nach welcher Definition auch immer) die meisten echten Fußgänger detektiert und nur relativ selten einen Falschalarm gibt. Solange — wie bei einem typischen Assistenzsystem — ein menschlicher Fahrer die Verantwortung trägt, kann das System so eingestellt werden, dass die Falschalarmrate sehr gering ist. Ein Notbremsassistent wird daher nicht laufend eine unnötige Vollbremsung durchführen. In den wenigen Fällen, in denen der Mensch kein Hindernis sieht, wird der Assistent zwar teilweise ebenfalls versagen und eben keine Bremsung einleiten. Das ist aber zu verschmerzen, denn eine Verbesserung stellt das System dennoch dar und die Verantwortung bleibt beim Mensch. Ein autonom handelndes Fahrzeug kann sich den Luxus dieser Parametrierung des Notbremssystems nicht leisten.

Schauen wir uns nun die Detektionsleistung einiger exemplarisch ausgewählter aktueller Forschungsarbeiten an. Dabei sind für uns „Detektionen" wichtig, d.h. es reicht nicht, dass ein Fußgänger im Bild gefunden wird, sondern es muss auch die Position einigermaßen gut (z.B. 50% Überlapp der umschließenden Rechtecke) gefunden werden.

Für die PASCAL Datenbank ergibt sich für die Kategorie „Person" mit dem besten untersuchten System eine sogenannte „Precision" von 95% ($0.95 = N_{korrPositiv}/(N_{korrPositiv} + N_{falschPositiv})$) bei einem sogenannten „Recall" von 0.9 [14]. D.h. wenn man das System so auslegt, dass in 95% der Detektionsfälle auch wirklich eine Person im Bild war, dann bleiben 10% der Bilder mit Personen unentdeckt. (Für andere Kategorien sind die Ergebnisse teilweise erheblich schlechter und die PASCAL Datenbank besteht aus vielen Einzelobjektbildern, d.h. vergleichsweise einfachen Szenen. Die COCO Datenbank liefert im Vergleich hierzu aufgrund der eher schwierigeren Szenen eine verminderte Detektionsleistung [13].

Zhang et al. haben für die (bereinigte) Caltech Datenbasis für die besten Algorithmen bei 0.1 falsch-positiven Detektionen pro Bild eine Fehlerrate (nicht-gefundene Fußgänger) von 10% gefunden [15]. Auf sehr ähnliche Ergebnisse kommen Li et al., die die Größenabhängigkeit der De-

tektionsleistung durch eine Kombination zweier Systeme in verschiedenen Skalen verbessern [16]. Andere Autoren (z.B. Wu et al. [17], Paisitkriangkrai [18] oder Tian et al. [19]) haben teilweise deutlich schlechtere Ergebnisse angegeben (bei gleicher falsch-positiv Rate).

Zusammenfassend kann man sagen, dass heute für gängige Datenbanken mit den besten publizierten Algorithmen in etwa 10% der Fußgänger nicht gefunden werden, wenn die Systeme so eingestellt werden, dass in jedem 10. Bild ohne Fußgänger ein scheinbarer Fußgänger detektiert wird. Die Ergebnisse der Detektion sind also vereinfacht ausgedrückt für 10% der Bilder falsch.

Das ist mit Sicherheit enttäuschend. Das Beispiel zeigt deutlich, wie groß der Unterschied zwischen Mensch und Maschine bei der Erkennung ist. Ich will dabei betonen, dass aus technischer Sicht die Ergebnisse durchaus beeindruckend sind und in den letzten Jahren hier extreme Fortschritte erzielt wurden[5]. Die Ergebnisse verdeutlichen daher nicht das Unvermögen der Entwickler sondern die Schwierigkeit der Aufgabe.

Glücklicherweise muss ein autonomes Fahrzeug nicht die volle Leistungsfähigkeit des menschlichen Sehsystems erzielen. Auch dies lässt sich am Beispiel der Fußgängerdetektion gut verdeutlichen. Sobald mittels der Gesamtsensorik (und Kartenmaterial) klar ist, wie ein Kamerabild einzuordnen ist, ist klar, wo überhaupt Fußgänger stehen könnten (nämlich im dreidimensionalen Raum auf dem Boden) und wie groß an dieser Stelle ein Fußgänger sein kann (Festlegung der Skalierung). Diese Zusatzinformation vereinfacht das Detektionsproblem erheblich und erlaubt so dann wesentlich bessere Ergebnisse für die Fußgängerdetektion. Auch die Kombination von verschiedenen Informationskanälen (Entfernung, Bewegung, Intensität) kann die Ergebnisse deutlich verbessern [20]. Einige Details zum Vorgehen für die öffentlichkeitswirksame autonomen Bertha-Benz Fahrt 2013 sind in [2] beschrieben. Ergebnisse zur Fußgängerdetektion für echte Szenen liegen mir nicht vor.[6]

Sobald deutliche Überdeckungen oder Verfremdungen der Objekte vorliegen sieht die Sache generell ebenfalls deutlich schwieriger aus. Je mehr verschiedene Perspektiven, Orientierungen (z.B. liegender Fußgänger statt stehender Fußgänger), Entfernungen ($\rightarrow$ Größe im Bild) auftreten, desto problematischer wird ganz generell die Detektion.

4.2 Ampeln, Hunde, Tüten und ihr Verhalten

Wie schwierig für die Maschine Objekterkennung ist zeigt sich im übrigen selbst bei viel weniger variablen Objekten. In [2] wird u.a. berichtet, dass selbst die bildbasierte Detektion von Ampeln alles andere als einfach ist, so dass eine Kartierung der Ampeln samt Speicherung entsprechender Bilddaten für jede Ampel notwendig wurde.

Es reicht aber nicht, Fußgänger und Ampeln zu detektieren. Betrachten Sie das von mir exemplarisch verwendete Kleinkind-auf-Bobbycar Beispiel. Für den Mensch liegt das Kleinkind (auf dem Bobbycar) zwar klar in der Klasse Mensch bzw. Fußgänger, aus Sicht des Computers wird aber vermutlich eher eine Einordnung in die Klasse „Hund" erfolgen. Und das ist nur ein Beispiel. Kartons müssen von Kisten unterschieden werden, im Wind flatternde Tüten von auf der Straße liegenden Menschen, Stahlbleche, die von Lastwägen fallen von leichten Styroporplatten. Schatten

[5] Die vor fünf Jahren von Dollar et al. publizierten Ergebnisse für verschiedene Datenbanken und Algorithmen lagen noch bei ca. 20% [11]. Es wurden also in der Tat erhebliche Fortschritte erzielt aber leider kann man hier nicht einfach extrapolieren, wie die Systemleistung nach weiteren fünf Jahren aussehen wird.

[6] Ergebnisse, die für eine reine Klassifikation von bereits ausgewählten Regionen gewonnen sind, sind mit kompletten Detektionen in Szenen nicht vergleichbar, siehe z.B. [11] für einen Vergleich.

von Löchern im Boden, Kühe von Werbplakaten etc. Tausende unterschiedlichster Objekte können für eine Entscheidung im Straßenverkehr relevant werden. Sicher, meistens, sind sie es nicht. Aber manchmal eben doch.

Der Mensch detektiert nicht nur, er segmentiert gleichzeitig auch in hoher Qualität und findet die Position im dreidimensionalen Raum, selbst anhand eines einzelnen zweidimensionalen Bildes. Vor allem aber detektiert er nicht nur die Hauptklasse sondern auch die Unterklassen (z.B. „Feuerwehrmann") und noch wichtiger: den Zustand des Objekts bzw. der Person.

Ein Beispiel soll dies verdeutlichen. Wenn Sie hinter einem Fahrradfahrer fahren, dann ist es für Sie durchaus im Sinne einer Vorhersage relevant, ob es sich z.B. um ein Grundschulkind (kennt Verkehrsregeln noch nicht und ist sehr spontan) oder einen Betrunkenen handelt. Sie werden entsprechend dieser Unterscheidung z.B. den mindestens einzuhaltenden Sicherheitsabstand unterschiedlich wählen. Auch Zusätze in der Szene können Einfluss haben. z.B. sehen Sie, dass der Fahrradfahrer gefährlich nahe parallel zu den Straßenbahnschienen fährt und daher mit nicht geringer Wahrscheinlichkeit demnächst stürzen wird. Oder Sie sehen, dass sein Fahrstil generell eher anarchisch ist und Sie mit sehr unerwarteten Fahrmanövern rechnen sollten.

Lange Rede kurzer Sinn: Selbst bei der absoluten Grundfähigkeit, der Objektdetektion versagt — im Vergleich zum Mensch — der Computer noch sehr deutlich und all die an sich wünschenswerten Zusatzinformationen, die der Mensch problemlos erfasst, können aktuell nicht detektiert werden.

5 Brauchen wir dieses Bildverstehen überhaupt?

Die Frage ist aber natürlich, ob es überhaupt notwendig ist, ein Bildverständnis zu erzielen, das ähnlich gut funktioniert, wie das des Menschen. In den meisten Fällen bleibt die noch nicht ausreichende Leistung glücklicherweise ohne Belang, denn in den meisten Situationen muss das Fahrzeug ja z.B. gar nicht alle Fußgänger detektieren. Wichtig sind nur die Fußgänger, die potentiell vor das Fahrzeug laufen und die Erkennung selbiger ist natürlich einfacher. Ein ausgefeiltes Bildverstehen ist also in der Regel nicht notwendig.

Die Wahrscheinlichkeit des Auftretens entsprechender kritischer Situationen lässt sich im übrigen natürlich zunächst massiv dadurch reduzieren, dass man das vollautonome Fahren nur unter eingeschränkten Szenarien zulässt. Wesentliche Vereinfachungen ergeben sich für die Sensorik, wenn starker Regen, Schnee und Nebel ausgeschlossen werden. Am schwierigsten sind sicher innerstädtische Fahrten und am einfachsten sind klar umrissene Situationen, wie das Fahren im Stau (welches wenig Anforderungen stellt und wenig riskant ist, so dass es bereits heute in verschiedenen Fahrzeugen angeboten werden kann). Die Fahrten auf Autobahnen sind zwar schnell (und damit potentiell gefährlich), aber — zumindest in Deutschland — ist der Zustand der Fahrbahn bzw. der Fahrbahnmarkierung sehr gut und es muss in der Regel nicht mit Fußgängern oder Fahrrädern gerechnet werden. Vieles was im Stadtverkehr problematisch ist (unvorhergesehene kleine Baustellen, Tiere, allerlei mögliche Objekte (von der Mülltonne bis zum Wahlplakat)) fallen bei der Autobahnfahrt weg. Allerdings: Sich darauf zu verlassen, dass all diese Dinge wegfallen kann riskant sein. Auch auf der Autobahn sind manchmal (wenn auch selten) spielende Kinder, Hunde, Vögel, Pferde etc. unterwegs. Dennoch ist die Auftretenswahrscheinlichkeit so gering, dass auch bei den begrenzten Mitteln heutiger Bildverarbeitung ein Betrieb eventuell möglich ist und weitere Einschränkungen (z.B. Fahrten nur auf speziell ausgewiesenen Autobahnen) sind denkbar.

Die zentrale Aussage an dieser Stelle ist aber: Auch wenn wir für die überwiegende Zahl an gefahrenen Kilometern nur ein rudimentäres Szenenverständnis und eine entsprechend einfache Objekterkennung benötigen, so gibt es doch immer wieder Fälle, in denen eine komplexe und schwierige Objekterkennung notwendig ist, um schwere Unfälle zu vermeiden. Selbst wenn dieser Fall nur einmal in einer Million Kilometern auftreten sollte bedeutet das, dass bei einer durchschnittlichen Jahresfahrleistung von 10.000 km über 10 Jahre jeder 10. Fahrer in einen entsprechenden Unfall verwickelt werden würde.

Leider können wir dies nicht in exakte Zahlen fassen. Dazu ist die Aufgabenstellung zu komplex. Wir wissen nicht, bei welchem Stand der Bildverarbeitung wieviele Kilometer gefahren werden müssen, um einen tödlichen Unfall zu provozieren. Von den Fahrzeugherstellern können Millionen von Kilometern als Testmaterial genutzt werden. Aber selbst bei einer Million gefahrener Kilometer: Wie oft tritt dabei ein brauner Bobbycar in Kombination mit einem Kleinkind und einem braunen Hund auf?

Die Kombinationsmöglichkeiten, die auftreten können sind astronomisch und nicht durch Standardtestfälle zu erfassen. Schon gar nicht sind genügend passende Testfälle vorhanden, um eine statistische Absicherung zu gewährleisten. Denn wenn wir z.B. für den Fall des braunen Bobbycars erzielen wollten, dass wir diesen mit 99% Wahrscheinlichkeit korrekt klassifizieren, dann würde das natürlich bedeuten, dass wir in der Größenordnung 10.000 passende Testbilder von braunen Bobbycars (in unterschiedlichsten Entfernungen, Größen, Ausgestaltungen etc.) vorliegen haben müssten.

Ich will im folgenden — recht willkürlich — einige mögliche Szenarien listen und der Leser möge sich jeweils selbst überlegen, an welcher Stelle das notwendige Bildverstehen wie technisch bewerkstelligt werden könnte.

S1 — Polizist, Anhalter oder Verletzer: Am Fahrbahnrand steht eine Person und winkt. Sollen Sie anhalten oder ist das nur ein freundliches Winken eine Kindes, ein unerwünschter Anhalter, ein Spaßvogel (schau mal wie einfach ich autonomes Fahrzeuge aus der Ruhe bringe) oder evtl. gar ein übel gesinnter Räuber? Genauso könnte es jemand sein, der mich vor einer Gefahr warnt oder Hilfe braucht.

S2 — Schwan voraus: Ein Schwan sitzt auf der Autobahn (zwischen zwei Fahrspuren) und der Verkehr zieht mit 100 km/h an ihm vorbei. Ist eine Vollbremsung angesagt? Können wir sicher sein, dass es ein Schwan ist?

S3 — fliegender Karton: Sie fahren auf der Autobahn in fließendem Verkehr hinter einem LKW und plötzlich segelt eine große Platte (ca. 2 x 1 m) von der Ladefläche. Handelt es sich um Styropor, leichte Kartonage, Holz oder Stahl? Machen Sie eine Vollbremsung, ein extremes Ausweichmannöver oder nehmen Sie den Aufprall in Kauf?

S4 — Kind zwischen Autos: Sie fahren in fließendem Stadtverkehr und sehen zwei Kinder teilweise sichtbar hinter parkenden Autos. Eine durchgezogene Linie trennt Sie von der Gegenfahrbahn. Ein Überfahren der Linie wäre eine Verletzung einer Verkehrsregel. Eventuell kommen zu allem Überfluss noch Fahrzeuge (wahlweise: LKW, Smart oder Fahrrad) auf der Gegenspur entgegen..

S5 — Gegenverkehr: Sie fahren auf einer Landstraße bei hohem Verkehrsaufkommen. Ein entgegenkommendes Fahrzeug startet ein Überholmannöver. Wie stark bremsen Sie ab? Versuchen Sie, die Fahrspur voll zu nutzen? Überfahren Sie die durchgezogene Linie? Fahren Sie rechts in die Wiese (kein Graben zwischen Wiese und Fahrbahn) ?

S6 — Umleitung: Sie fahren im fließenden Verkehr. Es hat sich ein Unfall ereignet, die Unfallstelle ist provisorisch abgesperrt und der Verkehr wird durch einen Polizisten umgeleitet. Können Sie (bzw. ihr Fahrzeug) die Gesten des Polizisten deuten?

S7 — Schlagloch: Ein großes Schlagloch (ca. 15 cm Tiefe) befindet sich vor dem Auto auf der Fahrbahn. Weichen Sie aus (Gefahr: Zusammenprall mit anderem Objekt bzw. Verletzung Verkehrsregel), fahren Sie weiter (Gefahr: Achsbruch) oder machen Sie eine Vollbremsung (Gefahr: provozierter Auffahrunfall)?

S8 — Plastiktüte: Eine große Einkaufstüte (oder war es doch eher ein Kind?) flattert über die Fahrbahn und Sie befinden sich in schnell fließendem Verkehr. Führen Sie eine Vollbremsung durch?

S9 — Parkende Autos: Offensichtlich (für Sie) parkt ein Auto in zweiter Reihe. Um das Auto „zu überholen" müsste die durchzogene Linie (Straßenmitte) überfahren werden. Das ist eine Verletzung einer Verkehrsregel. Werden Sie dies tun?

Wir haben hier teilweise schöne Beispiele, die zeigen, dass an sich vom autonomen Fahren profitiert werden kann. Das autonome Fahrzeug kann nämlich besser abschätzen, wieviel Platz jeweils rechts und links vom Auto vorhanden ist und so Ausweichmannöver sehr exakt durchführen. Auch der Bremsweg ist sehr gut berechenbar. D.h. sobald die Situation ausreichend gut erfasst ist, ergeben sich Vorteile für das autonome Fahrzeug. Andererseits ist natürlich die Grunderfassung schwierig und die Beispiele sind so gewählt, dass sie schwierige Entscheidungsfindungen beinhalten sobald eine Vollbremsung nicht mehr ausreichend ist. Für den Mensch ist es aufgrund des Flugverhaltens z.B. ein Leichtes, zwischen leichtem Karton und Stahl oder Schwan und Kind zu unterscheiden. Für das Fahrzeug allerdings nicht.

S5, eine an sich relativ klare Situation wird dadurch schwierig, dass es dem autonomen Fahrzeug unter Umständen schwer fallen wird, einzuschätzen, was die Fahrer der beiden anderen Fahrzeuge tun werden. Bremst der überholte Fahrer eher ab, weicht er nach links aus, wird er sich defensiv verhalten und bricht der Überholer noch ab, fällt er anderweitig negativ auf, wird er den Überholvorgang noch beschleunigen?

Beim aktuellen Stand wird gerne darauf verwiesen, dass ein Mensch ja im Fahrzeug vorhanden ist und gegebenenfalls die Steuerung übernehmen kann, denn verschiedenste Situationen erfordern heute definitiv ein Eingreifen des Menschen. Allerdings ist unklar, wieviele Sekunden hierfür veranschlagt werden müssen. Problematisch für den Gesamtverkehr (auch im Sinne von provozierten Unfällen) ist natürlich, wenn das Fahrzeug zunächst komplett abbremsen muss, bis der menschliche Fahrer die Situation erfasst und übernimmt.

6 Den Mensch kopieren

Wenn dieses Szenenverständnis so unermesslich schwierig ist, warum schlägt sich der Mensch darin so gut und wie können wir das nutzen, um entsprechende technische Systeme zu bauen?

Ein rein klassisches Bildverarbeitungsvorgehen, bei dem die Programmierer passende Detektoren für alle möglichen Objekte und Objektkombinationen und Szenen realisieren, ist bei der unendlichen Vielfalt relevanter Szenen und Objektkombinationen nicht realistisch. Ein Beispiel sind die oben beschriebenen Fußgängerdetektoren.

Es liegt nahe — analog zum Mensch —, auf lernende Ansätze zu setzen. Dementsprechend sind seit einiger Zeit wieder neuronale Netze in der Bildverarbeitung auf dem Vormarsch. Aktuell meist in der Form von sogenannten „Deep learning" Netzen [19, 21]. Beliebt (und auch sinnvoll) sind Netze mit verschiedenen Ebenen, die teilweise lokale Nachbarschaften ausnutzen (faltungsbasierte Schichten) [16]. Damit lassen sich die extremen Freiheitsgrade einschränken und somit die Lernzeit verkürzen. Zusätzlich werden hierdurch automatisch interne Kategorien und Vorverarbeitungen gebildet.

Können wir also nicht einfach ein neuronales Netz mit genügend (von Menschen beurteilten) Fußgänger- und Nichtfußgängerbildern damit trainieren? In einem gewissen Sinn ja. Aber unser Problem ist, dass wir zunächst genügend Bilder benötigen. Wir benötigen hierfür Millionen von Bildern, denn die Fußgänger können an ganz unterschiedlichen Positionen im Bild ganz unterschiedlich aussehen und orientiert sein. Je nach Entfernung und Alter sind sie unterschiedlich groß, vor allem müssen sie aber von allen möglichen Nicht-Fußgängerobjekten abgegrenzt sein. Wenn man andererseits zu viele Testbilder verwendet, dann kann es — je nach konkreter Netzausgestaltung — auch leicht zu einer Übergeneralisierung führen, d.h. die an sich wünschenswerte Generalisierungsfähigkeit des Netzes leidet, es werden dann also wirklich nur die vorgegebenen speziellen Fußgänger detektiert. Das Training eines neuronalen Netzes ist ein komplexes Unterfangen und die erzielte Leistungsfähigkeit hängt in komplexer Weise von allem ab (Lernalgorithmus, Testdaten, Netzstruktur, konkretes Neuronverhalten etc.) Wie gut kann man hierbei werden? Um es kurz zu machen: Gar nicht gut (s.o.).

Aber warum sind die Ergebnisse im Vergleich zum Menschen so schlecht? Letztlich weil auch hier keinerlei Gesamtverständnis für die Szene entsteht. Es wird eine isolierte Objekterkennung angestrebt und der Aufwand hierfür ist immens.

Sie mögen denken, dass reine Verbesserungen der Rechenleistung dazu führen, dass wir immer besser werden und sich die Leistung der Systeme schließlich dem Mensch annähert. Aber das ist fragwürdig.

Das menschliche Sehsystem ist mit Abstand das leistungsfähigste bildverarbeitende System das wir kennen. Das Sehen ist für den Mensch die zentrale Methode, um die Umwelt zu erfassen. Gegenüber anderen potentiellen Messmethoden erlaubt das Sehen eine Erfassung mit hoher Auflösung und Geschwindigkeit auch über weite Entfernungen. Es ist undenkbar, z.B. ein Pferd zu fangen, wenn man nichts sieht und etwa nur auf das Gehör angewiesen ist. Je ausgeklügelter die Erfassung der Umwelt ist, desto größer der Überlebensvorteil und so hat sich im Laufe der Evolution ein immer leistungsfähigeres Sehen entwickelt.

Wieviel Hirn für das Sehen notwendig ist, ist schwierig zu sagen, denn letztlich kann keine klare Grenze gezogen werden, was unabhängig vom Sehen ist und was nicht. Unterschiedliche Zahlen sind dementsprechend im Umlauf. Eine Annahme (s. [22]) geht davon aus, dass in etwa die Hälfte der menschlichen Großhirnrinde, die ca. 20 Milliarden Neuronen enthält, direkt oder indirekt mit dem Sehen beschäftigt ist. Wieviel Neuronen es genau sind ist für uns aber natürlich auch unerheblich. Wichtig ist stattdessen, sich klarzumachen, dass der Großteil unserer Informationsverarbeitungskapazität für das Sehen aufgewendet wird. Das Sehen ist unsere wichtigste Fähigkeit. Es

entscheidet über Leben und Tod. Und aus algorithmischer bzw. informationsverarbeitender Sicht ist es vermutlich auch unsere beeindruckendste Fähigkeit.

Man vergisst das leicht. Das Sehen scheint ja so einfach zu sein. Uns beeindruckt, wenn ein Computer den Weltmeister im Schach besiegt, meisterlich Jeopardy spielt oder Infektionskrankheiten besser als ein menschlicher Experte diagnostiziert. Wenn ein Schulkind aber in Sekundenbruchteilen eine Szene deutet, dann beeindruckt uns das nicht. Wir denken, dass das einfach sei und das Schachspielen schwierig. Natürlich ist es genau umgekehrt. Aber das fällt uns nicht auf, denn unser Gehirn wurde in hunderten Millionen Jahren nicht auf das Schachspielen optimiert, sondern auf das Sehen.

Wie sollten wir vorgehen, um die Leistungsfähigkeit des Gehirns zu kopieren? Wir könnten z.B. ein künstliches Gehirn bauen bzw. simulieren. Vermutlich ist es ausreichend, die Funktion der einzelnen Neuronen durch sehr idealisierte Schaltkreise (gewichtete Summation mit nachfolgender nichtlinearer Kennlinie) zu verwenden. Die eigentliche Leistungsfähigkeit liegt ja nicht in den Details der einzelnen Neuronen sondern in der Grundfunktion und der Verschaltung.

Bei 50% von 20 Milliarden Neuronen mit im Durchschnitt 10.000 Verknüpfungen (repräsentiert durch Gewichte) und einem Grundtakt von z.B. 20 Hz landen wir bei einer notwendigen Rechenleistung von „lediglich" $2 \cdot 10^{15}$ Multiply-Add Operationen, also 2000 TFlops. Eine moderne Grafikkarte bringt es auf 20 TFlops so dass wir festhalten können, dass die für das Sehen notwendige Rechenleistung in Zukunft technisch durchaus erreichbar sein sollte (auch wenn hier aktuell noch eine deutliche Lücke klafft). Und in der Tat ist ein großer Vorteil neuronaler Netze, dass sie — sobald die Gewichte der Neuronen erst korrekt gewählt, also gelernt, sind — vergleichsweise schnell Informationen verarbeiten und dabei auch noch sehr gut parallelisierbar sind.

Aber dennoch werden wir nicht so einfach zur ausgereift sehenden Maschine kommen. Das Problem liegt nämlich in der Bestimmung der $2 \cdot 10^{15}$ Gewichte. Diese müssen zunächst in irgendeiner Form erlernt werden. Und die hierfür notwendige Rechenleistung übersteigt alles, was wir uns vorstellen können.

Es ist nicht im geringsten so, dass der Mensch mit einem unprogrammierten neuronalen Netz im Kopf die Welt betritt. Natürlich lernen Menschenbabies und Kinder in den ersten Lebensjahren vieles und davon auch (in Kombination mit der Motorik) eine Menge was für das Sehen relevant ist. Die „Grundprogramme des Sehens", die diesen Lernvorgang des Individuums aber erst möglich machen, sind jedoch bereits im Erbgut des Menschen angelegt. Sie haben sich über die Jahrmillionen durch eine evolutionäre Optimierung ergeben [23].

Dabei reichen viele der elementaren Programme weit zurück. Katzen oder auch Hühnerküken unterliegen teilweise denselben optischen Täuschungen wie der Mensch (siehe z.B. Abb. 4, [24]); ein starkes Indiz dafür, dass die Grundfunktion des Sehens, die wir nutzen, weit zurück reichen, konkret bis zur Entstehung der Wirbeltiere. Trillionen von Individuen haben über hunderte Millionen von Jahren mittels genetischer Optimierung letztlich für die morphologische Grundstruktur unseres Gehirns und die Grundgewichte der Verbindungen gesorgt.[7]

Diesen Vorgang simulieren zu wollen ist schlicht undenkbar. Zum einen weil die dafür notwendige Rechenkapazität astronomisch wäre, zum anderen weil die entsprechenden Trainingsdaten nicht direkt nutzbar zur Verfügung stehen. Eine Messung der Morphologie und der Gewichte

[7] Man sollte sich das allerdings nicht so vorstellen, dass die Synapsengewichte direkt in der DNA in Form einer Eins-zu-Eins-Zuordnung kodiert sind. Vielmehr haben sich sehr effiziente genetische Programme gebildet, die bei der Morphogenese des Menschen ein sinnvoll arbeitendes Gehirn hervorbringen. Wie das im Detail funktioniert wissen wir nicht.

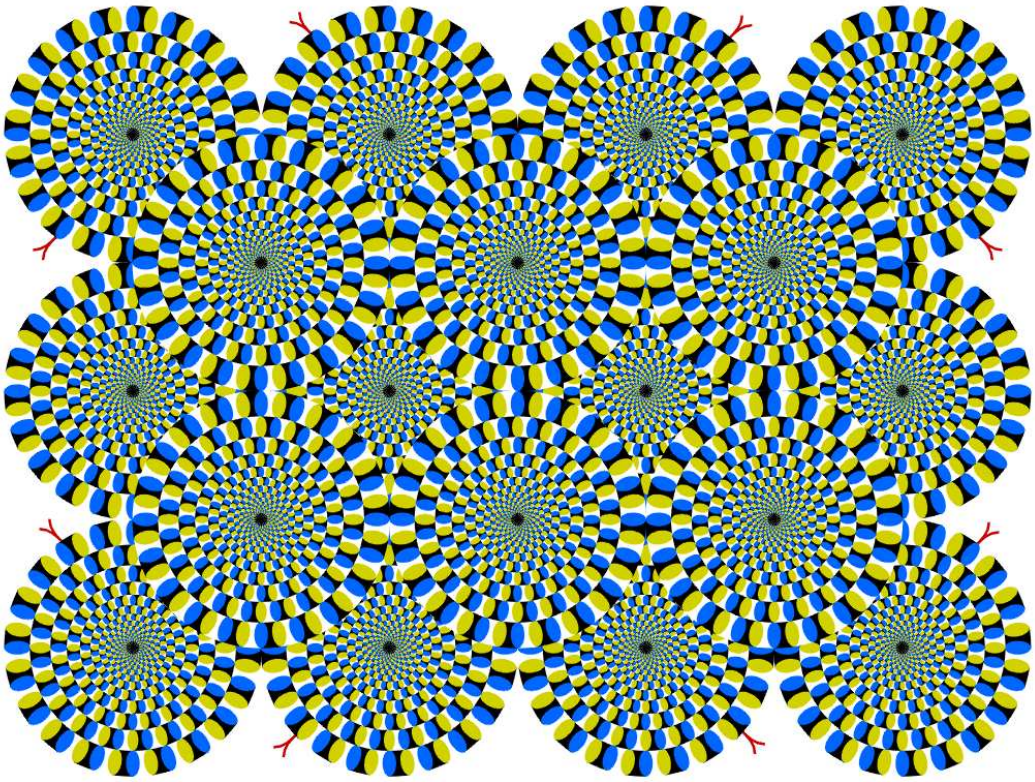
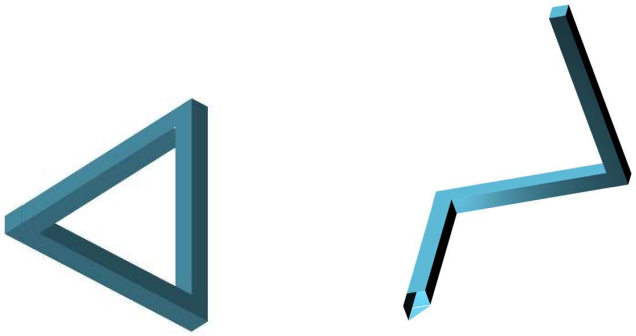


Abbildung 4 Periphere Drift Illusion: Beim Blinzeln oder anderen Augenbewegungen ergibt sich im peripheren Gesichtsfeld eine scheinbare Rotation der Kreise. Bild mit freundlicher Genehmigung von A.Kitaoka (Copyright September 2, 2003). Katzen unterliegen interessanterweise ebenfalls dieser Illusion (siehe z.B. https://www.youtube.com/watch?v=CcXXQ6GCUb8). Auch andere optische Täuschungen (siehe z.B. http://phys.org/news/2016-07-parrots.html oder [24]) werden von Tieren wahrgenommen. Dies ist ein starkes Indiz dafür, dass Teile unserer Sehprogramme sehr alt sind.

am menschlichen Gehirn (im Sinne eines Reverse Engineerings) ist ebenfalls beim Stand der Technik kein gangbarer Weg. Wir haben schlicht keine realistische Idee, wie das zu bewerkstelligen wäre.

Wir können auch nicht einfach einzelne Objekte, z.B. Tisch, dem System zeigen und damit im Sinne eines supervised Learning die Zuordnung zum Begriff „Tisch" angeben. Die Vielfalt der uns umgebenden Welt ist hierfür zu groß. Stattdessen nutzen wir auf allen Ebenen des Sehprozesses Zwischenergebnisse und „Regeln", die in den Gewichten der Neuronenverbindungen festgelegt sind. Die meisten dieser „Regeln" sind — vor allem auf tiefer, sensornaher Ebene — rein abstrakt. Höhere Regeln sind manchmal direkt benennbar. So scheint das Gehirn z.B. immer zu folgern, dass zwei sich im zweidimensionalen Netzhautbild schneidende Linien sich auch im dreidimensionalen Raum schneiden. Eine durchaus sinnvolle Regel, denn in den allermeisten Fällen führt sie bei dem Problem der Rückrechnung von zweidimensionalen Bildern auf die dreidimensionale Realität auf korrekte Resultate. Wir können sie künstlich verletzen und landen so bei einer großen Klasse optischer Täuschungen (ein bekanntes Beispiel zeigt Abb. 5).

Dies ist nur eine, für den Mensch direkt verständliche Regel. Unser Sehsystem ist voll von entsprechenden gelernten und in den Synapsen kodierten Regeln. Nochmals: „Gelernt" bedeutet dabei *nicht*, dass dieser Lernvorgang allein während der Lebensdauer des Individuums gelernt



(a) Ansicht 1 (b) Ansicht 2

Abbildung 5 Ein Beispiel für eine (im Laufe der Evolution) gelernte Regel Regel: Linien, die sich im zweidimensionalen Bild schneiden, schneiden sich (scheinbar bzw. in den allermeisten Fällen) auch in der dreidimensionalen Realität. Diese Regel führt dann zu verschiedenen optischen Täuschungen, z.B. dem „Penrose Dreieck": Aus der richtigen Richtung ergibt sich ein scheinbar unmögliches Dreieck. Wenn man die Beobachtungsrichtung ändert zeigt sich, dass in Wirklichkeit kein unmögliches Dreieck, sondern ein komplizierteres Objekt betrachtet wird.

wurde. Grundstrukturen, die für viele Muster und Programme notwendig sind, wurden im Laufe der Evolution entwickelt.

Zusammenfassend: Aktuell haben wir keine Ahnung, wie ein dem Mensch auch nur annähernd vergleichbares Sehsystem künstlich geschaffen werden könnte. Viele der in den letzten zehn Jahren erreichten Fortschritte im Bereich des maschinellen Sehens sind beeindruckend. Im Vergleich zur atemberaubenden Leistungsfähigkeit des menschlichen Sehsystems sind sie allerdings nichts.

7 Einige nichttechnische Probleme

Der Vollständigkeit halber will ich einige der nichttechnischen Probleme des autonomen Fahrens zumindest ansprechen, wenn mir auch die Kompetenz für eine Erörterung fehlt.

7.1 Ethische Probleme bei der Entscheidungsfindung

Es können — wenn auch selten — Situationen auftreten, bei denen eine Schädigung eines oder mehrerer Verkehrsteilnehmer nicht vermieden werden kann. In diesem Fall muss der Schaden minimiert werden. Wie ist aber der Schaden zu beurteilen?

Soll das Auto, wenn es einen von zwei Fahrradfahren treffen muss, eher den mit Helm (vielleicht überleben dann beide Fahrradfahrer) oder eher den ohne Helm (dann wird der Fahrradfahrer mit Helm nicht für sein Helmtragen bestraft) überfahren?

Wie groß muss die Wahrscheinlichkeit sein, dass das Objekt vor mir wirklich ein Kind ist, damit das Fahrzeug ausweicht und in die Hauswand rechts fährt (leichte Gefahr für Fahrzeugbesitzer und hohe Kosten am Fahrzeug und dem Haus)?

Solche Abwägungen sind unausweichlich und müssen getroffen werden. Natürlich trifft der Mensch diese Abwägungen auch — und zwar in Sekundenbruchteilen. Entscheidend ist aber, dass beim autonomen Fahrzeug diese Bewertungen und Maßstäbe fest vorgegeben werden. Durch wen? Und basierend auf welcher rechtlichen Grundlage?

7.2 Haftungsfragen

Die Gretchenfrage ist natürlich: Wer haftet im Fall eines Unfalls. Der Hersteller oder aber der menschliche „Fahrer", obwohl er überhaupt nicht fährt? Letzteres ist aus meiner Sicht für ein vollautonomes Fahrzeug kaum durchsetzbar und ich kann mir nicht denken, dass Kunden bereit sind, viel Geld für entsprechende Fahrzeuge auszugeben und dann schuldlos für hohe Schäden zu haften.

Rein finanzielle Schäden lassen sich im Prinzip mittels Versicherungen ausgleichen (und ich nehme an, dass in diesem Fall Versicherungen extrem teuer werden). Nicht-finanzielle Schäden sind aber kaum auf den nicht-fahrenden Kunden abzuwälzen.

Insgesamt, denke ich, dass hier keine wirklich befriedigende Lösung gefunden werden kann und — wegen der technischen Unlösbarkeit der Aufgabe — an dieser Stelle letztlich das Todesurteil für das autonome Fahrzeug gesprochen werden wird.

7.3 Wirkung auf den „Fahrer"

Gehen wir von dem Fall einer langen hochgradig (aber nicht vollständig) autonomen Autofahrt aus. Dies ist eine wichtige Zwischenetappe für viele Hersteller und den Gesetzgeber und wird in verschiedenen Kategorisierungen verwendet. Sie sollen als Fahrer also jederzeit das Steuer (und die Verantwortung) übernehmen können. Das wird Ihnen kaum gelingen, wenn sie während der Fahrt komplett entspannt sind oder sich einer anderen Tätigkeit (Schlafen, Fernsehen, Spielen ...) widmen. Wenn Sie stattdessen das Auto während der gesamten Fahrt kontrollieren, dann wird der Entspannungseffekt gering sein. Und: Je länger Sie fahren, ohne dass ein Eingreifen von Ihnen notwendig wird, desto mehr wird Ihre Aufmerksamkeit vermindert.

Ein weiterer langfristiger Effekt ist, dass die Fähigkeit des Autofahrens natürlich trainiert werden muss. Jemand der jahrelang nicht selbst fährt (sondern sich von seinem Auto fahren lässt), der wird Probleme haben, von jetzt auf gleich in einer Extremsituation das Steuer zu übernehmen und sich richtig zu verhalten.

7.4 Manipulation

Die möglichst gute Absicherung von Hard- und aber vor allem auch Software ist eine nicht zu unterschätzdende Schwierigkeit. Vom Terroranschlag über Mord bis zum einfachen Versicherungsbetrug sind viele unschöne Szenarien denkbar, wenn Unbefugte in der Lage sein sollten, elektronisch Einfluss auf das autonome Fahrzeug ausüben zu können oder gar komplett die Software ersetzen könnten. Entsprechende Hackerangriffe auszuschließen ist schwierig, ganz besonders vor dem Hintergrund, dass es jederzeit (zumindest durch Servicewerkstätten) möglich sein muss, die Fahrzeuge zu aktualisieren.

7.5 Kommunikation unter Menschen

Es ist uns gar nicht bewusst, wie oft wir im Straßenverkehr — meist sehr subtil und nicht direkt — mit anderen Fahrern oder Verkehrsteilnehmern kommunizieren. Ich bemerke, dass mich der andere Fahrer oder der Fußgänger gesehen hat und reagiere entsprechend. Dieses „Bemerken" ist oft nur ein kurzer Blickkontakt oder eine minimale Verlangsamung des Fahrzeugs oder ein leichter Richtungswechsel.

Auch bei Situationen mit unklarer Vorfahrtslage wird durch einfachste Kommunikation bzw. Deutung der Verhaltensweisen in aller Regel sehr schnell die Situation (z.B. wer fährt als erster) aufgelöst.

Im Sinne einer Car-2-Car Kommunikation sind entsprechende „Absprachen" natürlich noch einfacher realisierbar. Für gemischten Verkehr mit automatisierten und nicht-automatisierten Fahrzeugen ist die Situation schwierig.

8 Einwände

Für mich ist ziemlich klar, dass die Idee des vollautonomen Fahrens unter heutigen Verkehrssystemen in Stadt und Land, wenig sinnvoll ist und entweder zu einer massiven Verkehrsbehinderung (wenn die Systeme zur Unfallvermeidung extrem defensiv eingestellt sind) oder aber zu einem massiven Anstieg an Unfällen führen wird.

Aber natürlich sieht das nicht Jeder so. Die Thematik ist sehr komplex und es geht um die Projektion in die Zukunft. Insofern bleibt ein großer Spielraum für verschiedene Ansichten. Ich will zumindest auf einige der Standardargumente, mit denen Befürworter das autonomen Fahren oft anpreisen oder verteidigen, kurz eingehen.

Bedenken Sie beim Lesen entsprechender Experteninterviews, dass die üblicherweise befragten Experten in der Regel „Automobilexperten" oder Manager sind. Die von den an vorderster Front arbeitenden Ingenieure und Programmierer vertretenen Ansichten werden nur durch viele Managementebenen nach oben gereicht und gedeutet. Es ist mehr als fraglich, ob Vorstandsvorsitzende oder allgemeine Automobilexperten in der Lage sind, sich ein wirklich technisch fundiertes Bild bei diesem hochkomplexen Themenkomplex zu bilden.

A1: *„90% aller Unfälle lassen sich auf menschliche Fehler zurückführen.*" Das mag stimmen, verwundert aber nicht weiter, denn aktuell fährt ja auch der Mensch. Sobald die Mehrzahl der Autos autonom fährt, werden logischerweise die Mehrzahl der Unfälle von autonomen Fahrzeugen verursacht.

A2 *„Wir haben in den letzten Jahren so gewaltige Fortschritte gemacht und können heute bereits 99.99% aller Fahrsituationen beherrschen. In wenigen Jahren ist das autonome Fahren Realität.*" Eine typische falsche Extrapolation. 99.99% aller Fahrsituationen sind vergleichsweise einfach zu behandeln. Die fehlenden 0.01% sind aber nahezu unlösbar schwer. Der bisherige Ansatz des autonomen Fahrens wird diese Aufgaben (aufgrund des mangelnden Bildverstehens) nicht zufriedenstellend lösen können.

A3 *„Die Sensoren wurden in den letzten Jahren deutlich verbessert. Zukünftige Verbesserungen der Sensoren werden dafür sorgen, dass das autonome Fahrzeug wirklich autonom ist.*" Die Beschränkung des autonomen Fahrzeugs liegt nicht in der Sensorik sondern im Bildverstehen. Verbesserungen der Sensorik werden in der Tat autonome Fahrzeuge sicherer machen. Sie werden es aber nicht erreichen, die schwierigen Situationen, die ein echtes Bildverstehen erfordern, zu eliminieren.

A4 *„Wenn wir keine autonome Fahrzeuge bauen, dann werden wir von Apple, Tesla, Google und Co. überrollt und die Arbeitsplätze vieler Menschen in Deutschland sind in Gefahr.*" Angst als Argument ist immer sehr wirksam. Und in der Tat kann man — so man sich unsicher

ist, ob autonome Fahrzeuge funktionieren werden — einfach versuchen, an vorderster Front mitzumischen. Damit kann dieses (für Politiker und viele Menschen) Worst-Case-Szenario vermieden werden. Natürlich ist es aber kein echtes Argument hinsichtlich der Realisierbarkeit von autonomen Fahrzeugen sondern vielmehr ein Argument, warum weiterhin viel (Steuer-)Geld in das autonome Fahren investiert werden sollte.

A5 *„Daimler, Google, Apple etc. haben unendliche Geldmittel und die besten Ingenieure. Wenn die sagen, dass sie es schaffen, dann schaffen die das."* Absichtserklärungen und Wunschdenken führen nicht zwangsläufig zum gewünschten Erfolg. Die klare Vorgabe des Chefs, das dies und jenes zu erreichen ist, kann eine starke Fortschrittswirkung haben, kann aber auch bei unverwirklichbaren Zielen zu Problemen führen (vgl. VW-Abgasskandal). Vor allem aber ist anzuzweifeln, ob die entsprechenden Vorstände, Konzernstrategen, Entwicklungsleiter und Pressesprecher das notwendige Know-How haben und sich ausreichend von den „besten Ingenieuren" beraten lassen.

A6 *„Google Cars sind bereits Millionen von Kilometern in den USA vollautonom gefahren und haben bewiesen, dass die Technik funktioniert."* Leider ist die Wirklichkeit nicht so rosig, wie es in den Medien oft dargestellt wird [25, 26].

A7 *„Daimler hat mit der Bertha-Benz Fahrt bewiesen, dass ..."* Nein. Das ist kein Beweis. Zum einen muss man sehen, wie groß der Aufwand für diese spezielle Fahrt hierzu war. Zum anderen zeigt das veröffentlichte kurze Werbevideo natürlich nicht alle problematischen Fahrsituationen, bei denen der Betrachter sich eher fragt, was das Fahrzeug da um Himmels Willen gerade plant bzw. tut.

Vor allem kann aber eben anhand einer 100 km Fahrt keinerlei Aussage darüber getroffen werden, wie sich das Fahrzeug in seltenen Extremsituationen verhalten wird.

A8 *„Die Experten sind sich einig, dass das autonome Fahren kommen wird und bereits jetzt funktioniert."*

Nein. Die echten Experten, und das sind konkret Bildverarbeitungsexperten, sind sich darüber nicht einig.

A9 *„Die Technik für das autonome Fahrzeug ist fertig, lediglich die Gesetzeslage hindert uns daran, diese Fahrzeuge bereits heute zu verkaufen/nutzen."*

Nein. Wenn der Gesetzgeber die Fahrzeuge einfach so und ohne Beschränkung zulassen würde, dann hätten die Fahrzeughersteller ein großes Problem. Es würde sich nämlich sehr schnell zeigen, dass die Technik ganz und gar nicht fertig ist. Und natürlich stellt sich auch die grundlegende Frage, wer letztlich im Falle eines Unfalls haftet.

A10 *„Die Widerstände gegen das autonome Fahren entspringen der üblichen deutschen Technikfeindlichkeit."*

Nein. In der Tat gibt es Technikfeindlichkeit. Aber ein Hinterfragen der Technik ist unabdingbar und es hat sich in der Vergangenheit immer wieder gezeigt, dass ein blindes Vertrauen in den technischen Fortschritt und die diesen Fortschritt propagierende Industrie nicht unbedingt sinnvoll ist. Mit „Feindlichkeit" hat dieses kritische Hinterfragen nichts zu tun.

A11 *„Die üblichen Szenarien, die eine scheinbare Überforderung des autonomen Fahrzeugs zeigen, sind nicht praxisrelevant, denn die Fahrzeuge werden so defensiv und ggf. langsam fahren, dass nichts (schlimmes) passieren kann.*" In der Tat steigt die Unsicherheit jeder Fahrt nichtlinear mit der Geschwindigkeit. Ein Unfall bei 15 km/h hat in der Regel verkraftbare Folgen und die Bremswege und Reaktionszeiten sind unproblematisch. Auch stehende (weil von der Situation überforderte) autonome Fahrzeuge machen aktiv keinen Unfall. Allerdings führt ein solchermaßen extrem passives Fahrverhalten zu einem Zusammenbruch unseres Verkehrs und provoziert Unfälle anderer Verkehrsteilnehmer.

A12 *„Das autonome Fahrzeug ist gar nicht in der Lage, einen Gesetzesbruch durchzuführen, z.B. auf die Gegenfahrbahn zu fahren.*" Dies ist kein Feature sondern ein Fehler. Es gibt unzählige Situationen im Straßenverkehr, bei denen der menschliche Fahrer Verstöße gegen die Straßenverkehrsordnung in Kauf nehmen muss, um Schlimmeres zu verhindern oder auch nur einen Stau zu vermeiden. Das reicht vom Überfahren von durchgezogenen Linien bis über das Ausweichmanöver in den Graben bis zum Touchieren fremder Fahrzeuge (z.B. um zu vermeiden, dass der Fahrradfahrer auf der anderen Seite getroffen wird).

A13 *„Das autonome Fahrzeug muss nur so gut wie der Mensch werden.*"

Nein. Nehmen Sie an, dass ihr Kind a) von einem umsichtigen Autofahrer, b) einem betrunkenen Autofahrer oder c) von einem autonomen Fahrzeug tot gefahren wird. Ist das für Sie dasselbe? Wohl kaum. Obwohl in allen drei Fällen das Kind tot ist. Sie erwarten beim betrunkenen Autofahrer zu Recht eine erheblich höhere Strafe bzw. Sühne, denn der Fahrer hat bewusst in Kauf genommen, andere zu schädigen. Der Unfall war potentiell vorhersehbar. Wenn nun beim autonomen Fahrzeug klar ist, dass die Algorithmik so ausgelegt ist, dass in dieser speziellen Situation das Kind eben zwangsläufig überfahren wird (weil die Rechenleistung des preisgünstigen Fahrsystems eben nicht mehr hergibt), dann werden Sie es zu Recht nicht mit einer Entschuldigung der Form „tut uns leid, so ist das halt" bewenden lassen. Es nutzt im Einzelfall nichts, dass in Summe das autonome Fahrzeug „nur" genauso viele Fehler wie der Mensch macht. Zum einen werden es andere Fehler sein, zum anderen ist es ein Unterschied, ob ein Fehler — nach menschlichen Maßstäben — vermeidbar gewesen wäre oder nicht. Zumindest für die meisten Menschen.

A14 *„Wenn im Mittel die tödlichen Unfälle zurück gehen, dann sind autonome Fahrzeuge eine Erfolgsgeschichte.*" Nein. Zum einen kann der Verkehr nicht allein an Unfällen bzw. deren Vermeidung gemessen werden. Weitere Aspekte, z.B. Kosten aber auch Effizienz und durchschnittliche Transportzeit, sind ebenfalls von Belang. Zum anderen macht es für die meisten Menschen einen erheblichen Unterschied, ob ein Mensch bei einem Unfall ums Leben kommt weil ein anderer Mensch einen Fehler begangen hat, oder aber eine Maschine fest programmiert Fehlfunktionen aufweist. Nicht alle sehen das so und in der Tat besteht hier gesellschaftlicher Diskussionsbedarf.

A15 *„Daraus folgt nicht, dass das nicht funktioniert.*" In der Tat ist ein zukünftiges Nichtfunktionieren prinzipiell nicht zu beweisen, es können also nur Argumente angeführt werden, warum man es für wahrscheinlich hält. Bei dieser Argumentation schleicht sich allerdings auf gefährliche Weise eine Beweislastumkehr ein. Es ist nicht Aufgabe der Gesellschaft (oder

von Kritikern), das Nichtfunktionieren zu demonstrieren oder zu beweisen. Stattdessen ist es Aufgabe der Automobilkonzerne, das Funktionieren zu beweisen.

Nun wird natürlich nicht Jeder meiner negativen Einschätzung hinsichtlich der technischen Machbarkeit des vollautonomen Fahren zustimmen. Insbesondere die Entwicklungsabteilungen der Automobilkonzerne scheinen nach wie vor überzeugt zu sein, dass das schon alles ausreichend gut funktioniert. Ich würde mir auf folgende, an sich simple, Fragen (ich bleibe der Einfachheit halber beim „Bobbycar") Antwort wünschen. Und jeder Experte, der meint, sich sicher zu sein, dass autonome Fahrzeuge dem Mensch hinsichtlich Sicherheit überlegen sind, sollte sich diese Fragen selbst ehrlich und überzeugt beantworten können.

1. Bei der Fahrt entlang parkender Autos: Welche maximale Geschwindigkeit werden sie zulassen und welcher Bremsweg ergibt sich?

2. Annahme: in 5 m Entfernung rennen bzw. rollen ein Hund (von rechts) und ein Kleinkind auf braunem Bobbycar (von links) auf die Straße. Sie können nicht beiden gleichzeitig ausweichen. Sehen Sie es also als notwendig an, zwischen braunem Bobbycar mit Kleinkind und Hund zu unterscheiden?

3. Kann das Fahrzeug das heute schon (oder bald) unterscheiden? Wenn ja, wie macht es das technisch? Was für eine Art von neuronalem Netz wird verwendet? Faltungsbasierte Schichten? Wieviele? Polling? etc.,). (Ein Experte, der behauptet, dass das technisch funktioniert, sollte diese Fragen beantworten können.)

4. Mit welchem Bildmaterial erfolgte der statistische Nachweis? Wieviele passende Bobbycar Bilder wurden beim Lernen und Testen verwendet? Wie wurden die generiert? Wie sieht es mit den Hunden aus? Welche Fehlalarmrate und welche Erkennungsrate haben Sie damit erzielt?

5. Wo kann man das nachlesen?

Kann man die Antwort auf all diese Fragen mit einem lapidaren Hinweis auf Geheimhaltung verweigern? Auch bei der Zulassung eines neuen Medikaments erwarte ich vom Hersteller, dass von ihm klar dokumentiert und veröffentlicht ist, wie die Wirksamkeit und die potentiellen Gefahren untersucht wurden und welches Ergebnis dabei erzielt wurde. Beim autonomen Fahren geht es ebenfalls um Leben und Tod und daher ist es auch dort nicht zuviel verlangt, entsprechende Nachweise zu erbringen.

9 Experten und Öffentlichkeit

Warum sind sich trotz der offensichtlichen Schwierigkeiten scheinbar alle Experten sowie die Öffentlichkeit (vertreten durch die Medien) einig, dass das autonome Fahren in wenigen Jahren Realität wird?

Natürlich wird verborgen in all den optimistischen Einschätzungen — selbst von den Herstellern — teilweise auf die Schwierigkeiten verwiesen. So schreiben Aeberhard et al. in [1] z.B. „Simple situations can be easily modeled, but there are always exceptions and unexpected situations that can happen at any time; developing the algorithms to react correctly in these unique

situations is still quite challenging. The challenges in artifical intelligence for automated driving systems will always have their limits, but will also continously improve until a level of intelligence is reached with which highly automated driving will be possible and where safety, within certain conditions, can be guaranteed."

Man beachte, dass hier explizit auf das "highly automated driving" (gegenüber dem voallautonomen Fahren) eingeschränkt wird. Ganz generell beherrscht aber der Optimismus die öffentliche Meinung. Herfür sind mehrere Gründe ausschlaggebend:

Komplexität: Die Thematik ist komplex und es ist nicht trivial, die Behauptung zu überprüfen, dass autonomes Fahren bald funktionieren wird. Als auf die Zukunft bezogene Aussage ist sie natürlich per se nicht falsifizierbar, aber es fällt auch nicht leicht, die Wahrscheinlichkeit für das Eintreffen abzuschätzen.

Autoritäten: Das Vertrauen in scheinbare Autoritäten auf einem Gebiet (z.B. Daimler (Auto), Google (Informationsverarbeitung), Bundesverkehrsministerium (Seriosität), Presse (Glaubwürdigkeit)) ist generell hoch und in der Regel ist dieses Vertrauen auch sinnvoll. Wenn ich keine anderen Anhaltspunkte habe, dann ist es das Sinnvollste, sich der Meinung von Experten anzuschließen (wenn es denn Experten sind).

Übernahme von Pressemeldungen: In der heutigen Zeit verbreiten sich berichtenswerte Nachrichten — ob richtig oder falsch — in Windeseile. Meldungen einer Nachrichtenagentur werden von unzähligen Blogs, RSS feeds und anderen Medien aufgegriffen und nacherzählt. Das Ergebnis ist, dass eine Aussage oft innerhalb weniger Tage auf hunderten oder gar tausenden von Informationskanälen verfügbar ist. Sobald jemand nach passenden Informationen sucht, wird er letztlich immer wieder bei derselben Grundaussage (z.B. „Google Autos sind 1 Million km autonom gefahren") landen.

Fortschrittsglaube und falsche Extrapolation: Aus den Erfolgen der letzten Jahren wird extrapoliert, dass das Problem des autonomen Fahrens bald gelöst ist. Leider ist die hier zu extrapolierende Funktion (z.B. beherrschte Fahrsituationen im Laufe der Zeit) extrem nichtlinear. Das führt dazu, dass die Extrapolation letztlich versagt.

Experten: Die meisten Experten sind nicht wirklich Experten hinsichtlich der Bildverarbeitung oder dem Sehen.

Experten und Experten: Experten haben die starke Tendenz sich der Meinung anderer Experten anzuschließen.

Eigeninteresse: Einige Experten auf dem Gebiet profitieren natürlich massiv von der Entwicklung bzw. den in diesen Bereich gesteckten Finanzmitteln (seien es Steuergelder oder auch Industriemittel). Diese Eigeninteressen müssen nicht mal zu bewussten Falschaussagen führen, können aber dennoch unterbewusst die Meinung des Experten beeinflussen.

Es sei betont, dass hier nirgends unlautere Absichten zu unterstellen sind. Diese Prozesse der Meinungsbildung verlaufen so wie sie verlaufen.

10 Folgerungen für die Gesetzgebung und den Staat

Das komplette Funktionieren des autonomen Fahrens zu beweisen ist per se unmöglich und dementsprechend auch nicht einzufordern. Es ist allerdings zu fordern, dass dieser Nachweis so gut wie irgend möglich erfolgt. Hierzu müssen die genauen Tests und statistischen Annahmen klar dargestellt werden. Die Zusammenhänge sind sehr kompliziert weil es letztlich unendlich viele mögliche Fahrsituationen bzw. Szenen gibt. Umso wichtiger ist hier ein offener und klarer Umgang mit den eingesetzten Verfahren zur Bildverarbeitung und zur statistischen Absicherung.

Es ist verständlich, dass die Autoindustrie daran wenig Interesse hat und die Geheimhaltung als essentiell ansieht. Damit kann man aber letztlich nicht zufrieden sein. Auch bisher wird sorgfältig entschieden, ob Fahrzeuge sicher sind und dementsprechend zugelassen werden. Diese Prüfung muss unabhängig von den Herstellern durchgeführt werden. Dies ist ein sehr problematischer Punkt, denn natürlich ist sie extrem aufwändig und eigentlich nur annähernd zu bewerkstelligen, wenn die Details der eingesetzten Sensoren und Algorithmen offengelegt werden. Hierzu genügend unabhängige Expertise an staatlichen Stellen aufzubauen wird ein schwieriges Unterfangen werden.

Die Kernkomponente des autonomen Fahrens ist das künstliche bzw. maschinelle Sehen (sei es mit konventionellen 2D Abbildungen, Lidar oder Radar). Zum „Sehen" gehört hier die gesamte Wahrnehmungskette von der Sensorik bis zur kompletten Repräsentierung der Realität in einem Modell (unterstützt durch Vorwissen, z.B. Kartenmaterial oder anderes „Wissen").

Ich habe dargestellt, warum wir aus meiner Sicht die menschliche Leistung in diesem Teilbereich der „Intelligenz" nicht in irgendeiner absehbarer Zukunft erreichen werden und dementsprechend die Verheißungen der Autonomes-Fahren-Lobby nicht wirklich erfüllbar sind.

Dennoch sind die eingesetzten Steuermittel nicht verschwendet. Die Forschung in diesem Bereich ist im Sinne des Erkenntnisgewinns und der Grundlagenforschung wichtig und eine Vielzahl von Anwendungen fernab vom autonomen Fahren würde von verbessertem maschinellen Sehen profitieren. Im Sinne der Ehrlichkeit sollte allerdings ein realistisches Bild vom autonomen Fahren und seinen Problemen vermittelt werden.

Selbst wenn wir nicht daran glauben, ähnlich gut wie der Mensch zu werden, bedeutet das natürlich nicht, dass wir den Traum von sehenden Maschinen aufgeben sollten. Es ist aus meiner Sicht der wichtigste und spannendste Bereich der sogenannten künstlichen Intelligenz und — wie bereits dargelegt — der Bereich, den der Mensch aufgrund des damit einhergehenden Überlebensvorteils perfektioniert hat.

Dass im Bereich des autonomen Fahrens/Fahrerassistenz aktuell das Hauptinteresse besteht, liegt schlicht daran, dass es sich hierbei um ein potentiell riesiges Marktvolumen handelt, das insbesondere für die deutsche Wirtschaft von enormer Bedeutung ist. Sowohl staatliche als auch industrielle Forschungsgelder sind daher verhältnismäßig einfach verfügbar.

Die Anwendungsmöglichkeiten von sehenden künstlichen Systemen sind dagegen unvorstellbar. Sie reichen von der Überwachung von Kranken, über die automatisierte Diagnostik von Krebs über die verbesserte und verbilligte Produktion bis zum Bereich der Verbrechensprävention. Und natürlich ist auch die Fahrerassistenz eine unter tausender wichtiger Anwendungen.

Nahezu überall wo traditionell Menschen eingesetzt werden, wird das Sehen des Menschen als Hauptsensorik genutzt. Von der Montage von Handys über die Pflege von Kranken bis zum Verkauf von Backwaren. Sehen ist immer notwendig. Leistungsfähiges künstliches Sehen kann dementsprechend zu einer Rationalisierung in ungeahntem Ausmaß führen. Und das bedeutet

positive wie negative Folgen, kann insbesondere aber auch mittelfristig eine Umgestaltung des Gesellschaftssystems notwendig machen.

Und eine weitere weitreichende Auswirkung sollte aus meiner Sicht ebenfalls bedacht werden: Die Technik des autonomen Fahrens ist natürlich (und sehr viel einfacher als beim konventionellen Automobil) in Waffensysteme direkt übernehmbar. Von Drohnen bis zu Kampfrobotern, autonome Waffen werden am direktesten von den Fortschritten beim künstlichen Sehen profitieren. Auch hier sind die potentiellen Folgen für die Gesellschaft und die Welt mittelfristig vermutlich dramatisch, denn es steht zu befürchten, dass zur Konfliktlösung schneller Kriege eingesetzt werden könnten.

Angesichts dieser Problematik beschleicht mich doch ein ungutes Gefühl und die Frage stellt sich, ob die Büchse der Pandora nicht besser geschlossen bleiben sollte. Man mag das als typisch deutsche Technikfeindlichkeit oder Bedenkentum abtun. Aber: Eine Diskussion dieser Aspekte ist — genau wie z.B. im Bereich der Gentechnik — notwendig und unverzichtbar.

Sollten wir als Steuerzahler also weiter Geld in diesen Bereich investieren und sollten wir als Ingenieure weiter unser Herzblut in das künstliche Sehen stecken?

11 Zusammenfassung

Wagen wir einen Blick in die Zukunft. Werden wir in den nächsten zehn Jahren autonom fahren? Ganz unterschiedliche Einschätzungen sind möglich.

E1: Das autonome Fahrzeug wird technisch realisiert und fährt mindestens so gut wie der Mensch. Autonomes Fahren wird Realität.

E2: Es ergeben sich zwar technische Defizite, aber letztlich wird unser Verkehrssystem so umgebaut, dass die Restrisiken tragbar sind. Autonomes Fahren wird Realität.

E3: Es ergeben sich zwar technische Defizite, aber dennoch wird das autonome Fahren (zu Lasten der Sicherheit und der Effizienz) großflächig eingeführt.

E4: Autonome Fahrzeuge werden verfügbar, aber werden nur von wenigen Menschen gekauft, da die Haftung auf den Fahrer abgewälzt wird.

E5: Die Defizite des autonomen Fahrens in Kombination mit der Haftungsproblematik führen dazu, dass das Vorhaben letztlich aufgegeben wird. Die Verantwortung bleibt bei dem durch ausgefeilte Fahrerassistenzsystemen unterstützten menschlichen Fahrer.

Ich vertrete Einschätzung E5 und habe in dieser Darstellung begründet, warum aus meiner Sicht vollautonomes Fahren im Kontext des heutigen Verkehrs in einer dem Mensch vergleichbaren Qualität nicht in absehbarer Zeit erzielbar sein wird. Das Grundproblem ist, dass zwar in verschiedenen Fahrsituationen das autonome Fahrzeug potentiell sicherer fahren kann, dass aber in einem nicht-vernachlässigbaren Anteil von Fahrsituationen ein komplexes Bildverstehen notwendig sein wird. Beim Stand der Technik ist die technische Realisierung eines entsprechenden Bildverstehens nicht möglich.

Assistenzfunktionen werden weiter zunehmen. Problematisch wird bereits der Schritt zum teilautonomen Fahren werden. Wenn das Auto in 99% der Fälle keinen Eingriff des Fahrers benötigt, dann wird es für den Fahrer schwierig werden, die Aufmerksamkeit so hoch zu halten, dass er jederzeit eingreifen kann. In speziellen Fahrsituationen ist der Einsatz des autonomen

Fahrens ohne Risiko oder mit sehr begrenztem Risiko bereits heute technisch möglich. Ein Beispiel ist die Fahrt in einem Stau. Auch Fahrten auf Autobahnen sind denkbar (wenn auch hier das Risiko für einen tödlichen Unfall aufgrund der höheren Geschwindigkeit natürlich deutlich steigt). Vermutlich werden also entsprechende autonome Fahrfunktionen für bestimmte Verkehrssituationen kommen. Der Versuch, die sich ergebende Fortschritte auf das allgemeine Fahren in Stadt und Land zu übertragen wird dagegen (mehr oder weniger kläglich) scheitern.

Abschließend möchte ich den verschiedenen Lesertypen dieser kurzen Abhandlung ungefragt jeweils einen Ratschlag mit auf den Weg geben.

Dem fachfremden Laien: Das menschliche Sehsystem ist unerreicht. Erstarren Sie in Ehrfurcht vor dieser unserer beeindruckendsten informationsverarbeitenden Leistung und machen Sie sich klar, warum und wie sich dieses Sehen gebildet hat.

Dem Politiker: Lassen Sie sich nicht von Floskeln und selbstbewussten Behauptungen täuschen sondern hinterfragen Sie und bekommen Sie ein Gefühl für die technischen Grenzen.

Dem jungen Ingenieur: Überdenken Sie, wie und wo Ihre Arbeit eingesetzt werden wird.

Dem Manager: Dieses Thema ist — im Gegensatz zur Rationalisierung bei der Produktion von Kaffeetassen — nicht für eine Politik von oben im Sinne von Zielvorgaben geeignet. Hören Sie auf ihre Ingenieure und lassen sie diese ein *realistisches* Bild der Möglichkeiten zeigen.

Literatur

[1] M. Aeberhard, S. Rauch, M. Bahram, G. Tanzmeister, J. Thomas, Y. Pilat, F. Homm, W. Huber, and N. Kaempchen, "Experience, results and lessons learned from automated driving on germany's highways," *IEEE Intelligent Transportation Systems Magazine* **7**(1), 42–57 (2015).

[2] T. Dang, M. Lauer, P. Bender, M. Schreiber, J. Ziegler, U. Franke, H. Fritz, T. Strauß, H. Lategahn, C. G. Keller, *et al.*, "Autonomes fahren auf der histlorischen bertha-benz-route," *tm-Technisches Messen* **82**(5), 280–297 (2015).

[3] J. L. Miller, P. Clayton, and S. F. Olsson, "Overview of benefits, challenges, and requirements of wheeled-vehicle mounted infrared sensors," in *SPIE Defense, Security, and Sensing*, 87040I–87040I, International Society for Optics and Photonics (2013).

[4] B. Tippetts, D. J. Lee, K. Lillywhite, and J. Archibald, "Review of stereo vision algorithms and their suitability for resource-limited systems," *Journal of Real-Time Image Processing* **11**(1), 5–25 (2016).

[5] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, **2**, 807–814, IEEE (2005).

[6] "Middlebury Stereo Evaluation - Version 2 ." `http://vision.middlebury.edu/stereo/eval/`. Accessed: 2016-08-22.

[7] M. Pijpers, "Capita selecta: Virtual reality," *Sensors in ADAS, A literature study on the working principles and characteristics of frequently applied sensors in Advanced Driver Assistance Systems* **8** (2007).

[8] "Continental / A.D.C GmbH, ARS 408-21 Premium Long Range Radar Sensor." http://www.conti-online.com/www/download/industrial_sensors_de_de/themes/download/ars_premium_datenblatt_de.pdf. Accessed: 2016-08-22.

[9] C. Beder, B. Bartczak, and R. Koch, "A comparison of pmd-cameras and stereo-vision for the task of surface reconstruction using patchlets," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8, IEEE (2007).

[10] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE Journal of quantum electronics* **37**(3), 390–397 (2001).

[11] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE transactions on pattern analysis and machine intelligence* **34**(4), 743–761 (2012).

[12] J. Tao, M. Enzweiler, U. Franke, D. Pfeiffer, and R. Klette, "What is in front? multiple-object detection and tracking with dynamic occlusion handling," in *International Conference on Computer Analysis of Images and Patterns*, 14–26, Springer (2015).

[13] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European Conference on Computer Vision*, 740–755, Springer (2014).

[14] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision* **88**(2), 303–338 (2010).

[15] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele, "How far are we from solving pedestrian detection?," *arXiv preprint arXiv:1602.01237* (2016).

[16] J. Li, X. Liang, S. Shen, T. Xu, and S. Yan, "Scale-aware fast r-cnn for pedestrian detection," *arXiv preprint arXiv:1510.08160* (2015).

[17] S. Wu, R. Laganière, and P. Payeur, "Improving pedestrian detection with selective gradient self-similarity feature," *Pattern Recognition* **48**(8), 2364–2376 (2015).

[18] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Pedestrian detection with spatially pooled features and structured ensemble learning," *IEEE transactions on pattern analysis and machine intelligence* **38**(6), 1243–1257 (2016).

[19] Y. Tian, P. Luo, X. Wang, and X. Tang, "Pedestrian detection aided by deep learning semantic tasks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5079–5087 (2015).

[20] M. Enzweiler and D. M. Gavrila, "A multilevel mixture-of-experts framework for pedestrian classification," *IEEE Transactions on Image Processing* **20**(10), 2967–2979 (2011).

[21] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 3642–3649, IEEE (2012).

[22] S. Hagen, "The mind's eye," *Rochester Review Marc* **March 2012**, 32–37 (2012).

[23] E. B. Baum, *What is thought?*, MIT press (2004).

[24] O. Rosa Salva, R. Rugani, A. Cavazzana, L. Regolin, and G. Vallortigara, "Perception of the ebbinghaus illusion in four-day-old domestic chicks (gallus gallus)," *Animal Cognition* **16**(6), 895–906 (2013).

[25] "Google Self-driving Car testing report on disengagements of autonomous mode December 2015 ." `https://www.google.com/search?hl=en&q=SDC+disengagements+Report&gws_rd=ssl`. Accessed: 2016-08-22.

[26] "6 Simple Things Google's Self-Driving Car Still Can't Handle." `http://gizmodo.com/6-simple-things-googles-self-driving-car-still-cant-han-1628040470`. Accessed: 2016-08-22.