

Das Automobil als Sehmaschine

Fabian Kröger

Über den größten Zeitraum seiner Geschichte hinweg war das Automobil blind. Es konnte seinen Weg nicht allein nach Hause finden, wie es die Pferde noch schafften, wenn der Kutscher eingeschlafen war. Das Automobil war auf einen menschlichen Chauffeur angewiesen, der die Umgebung des Fahrzeugs mit seinem Sehsinn ständig im Auge behalten musste. Um das Fahren auch nach Einbruch der Dunkelheit zu ermöglichen, wurde das Auto bald mit Acetylenlampen ausgerüstet. Da ihr Licht aber nur aus kurzen Distanzen sichtbar war und keine große Reichweite hatte, wuchs der Wunsch nach einer erhöhten Sichtbarkeit des Autos, aber auch der Umgebung. In den 1920er Jahren wandelte sich die Beleuchtung von einem »passiven Mittel des Gesehenwerdens«¹ zum voraustastenden elektrischen Scheinwerfer. Er machte das Automobil nicht nur unabhängig von Witterung und Tageszeiten, sondern ermöglichte auch höhere Geschwindigkeiten. Scheinwerfer waren die ersten vorausschauenden Augen des Automobils, blieben aber auf die menschlichen Blicke bezogen, die aus dem Inneren heraus und von außen auf das Fahrzeug geworfen wurden.

Mit der in den 1950er Jahren begonnenen Forschung zum automatisierten Fahren wurde die Entwicklung eines neuen, sensorischen Auges notwendig: Es musste ein Weg gefunden werden, das Auto – nun ohne den Umweg über humane Perzeptionsfähigkeiten – mit seiner Umgebung zu verschalten. Um digitale und analoge Räume in *real-time* zu synchronisieren, musste eine zuverlässige Verbindung, ein Kommunikationsprozess zwischen Automobil und Umwelt organisiert und die Welt in eine maschinenlesbare Form übersetzt werden.² Diese Mediation zwischen physikalischer Außen- und digitaler In-

1 Möser, Kurt: Geschichte des Autos, Frankfurt a.M.: Campus 2002, S. 45.

2 Vgl. Gugerli, David: Wie die Welt in den Computer kam. Zur Entstehung digitaler Wirklichkeit, Frankfurt a.M.: S. Fischer 2018, S. 194.

nenwelt wird durch Sensoren bewerkstelligt, also Vorrichtungen, die etwas messen oder detektieren und daraus ein maschinenlesbares Signal erzeugen, das von einem Rechner verarbeitet werden kann.

Zwar ist das automatisierte oder sogar autonome Fahren bis heute keine Realität geworden, unsere noch immer primär manuell gesteuerten Fahrzeuge sind aber seit Beginn der 1990er Jahre zu Sehmaschinen geworden – Kamera-, Radar-, Ultraschall-, Infrarot- und Lidarsensoren werden von *Advanced Driver Assistance Systems* (ADAS) genutzt, die den Fahrvorgang inkrementell automatisieren. Während automatische Fahrzeuge für bestimmte Aufgaben programmiert sind und somit ein vorhersehbares Verhalten haben, gehen autonome Systeme darüber hinaus, da sie selbst Entscheidungen – etwa für einen Spurwechsel oder ein Bremsmanöver – treffen. Sehende Fahrzeuge sind also nicht unbedingt autonom, autonome Fahrzeuge sind aber zwingend Sehmaschinen. Es ist denkbar, dass die immer weiter entwickelten Wahrnehmungsfähigkeiten von Autos in Zukunft zu einer schrittweisen Einführung des autonomen Fahrens führen werden.

Der folgende Beitrag befasst sich mit dem Wahrnehmungsvermögen der oben genannten Fahrzeugsensoren. Dazu ist es nützlich, ihre Entwicklung in einem größeren technik- und kulturhistorischen Kontext zu situieren. Der Medienwissenschaftler Simone Arcagni vertritt die These, dass jede historische Periode von einer spezifischen Visualität, einem ›Auge‹ geprägt ist. Damit meint er eine bestimmte Mentalität und Sehlogik, die ihren Ursprung in der Gesellschaft habe und diese Gesellschaft zugleich forme.³ Mit dem Aufkommen von Photographie und Kino im 19. Jahrhundert sei das linsenbasierte mechanische Auge der Kamera zum dominanten Modell geworden. Mit der Computertechnologie der zweiten Hälfte des 20. Jahrhunderts habe dann aber ein neues, mathematisches oder kybernetisches Auge die Bühne betreten, das auch mit dem Begriff *Machine Vision* bezeichnet werden kann.

Arcagni betont, dass dieses kybernetische Auge auf völlig andere Weise funktioniere als die optische Kamera des 19. Jahrhunderts. Es gewinne seine Informationen nicht mit optischen, sondern mathematischen Instrumenten, mit Codes und Algorithmen. Auch mit dem Gesichtssinn kann dieses neue Auge nicht direkt verglichen werden: Der französische Kulturtheoretiker Paul Virilio (1932–2018), der sich in seinen Essays mit dem Zusammenhang von Krieg, Medien und Geschwindigkeit befasste, wies schon Mitte der 1980er Jahre darauf hin, dass das Auge der *Machine Vision* nicht nur eine mehr oder

3 Arcagni, Simone: *L'occhio della macchina*, Turin: Einaudi 2018.

weniger elaborierte Imitation des menschlichen Auges ist, sondern seine Ersetzung durch ein anderes Auge vornimmt.⁴

Um was für ein Auge es sich hier handelt und wodurch es sich vom optischen, aber auch menschlichen Auge unterscheidet, soll im Folgenden erläutert werden. Ziel ist es, das Zusammenspiel von Sensoren, dem maschinellen Sehen experimenteller Blicksysteme und den daraus resultierenden Bildern sowie ihrer Verarbeitung im Rechner im Hinblick auf zukünftige autonome Fahrzeuge zu begreifen. Dabei soll gezeigt werden, dass der Blick ein produktiver theoretischer Zugang zur *Machine Vision* sein kann. Methodisch wird dabei der Versuch unternommen, Erkenntnisse der Kultur- und Medienwissenschaften – insbesondere der *Visual Culture Studies* – mit einem technikhistorischen Blick auf Entwicklungen der Ingenieurwissenschaften in einen Dialog zu bringen. Da das Nachdenken über das Auge der Fotografie, den Blick von Apparaten und die von ihnen produzierten Bilder zu den Kernkompetenzen dieser Fächer gehören, bietet es sich an, die Entwicklung von autonomen Sehmaschinen mit diesen Disziplinen einzuordnen. Konkret geschieht dies durch eine Verschränkung der Thesen Paul Virilios mit der Entwicklung eines sehenden Automobils in einem vom Ingenieur Ernst Dieter Dickmanns geleiteten Forschungsprojekt.

Im ersten Abschnitt sollen die verschiedenen Sensoren vorgestellt werden, die autonome Fahrzeuge benötigen, um die Umgebung wahrzunehmen und ohne Fahrer:in zu navigieren. Im zweiten Teil soll einer dieser Sensoren, die Kamera, genauer untersucht werden. Dazu wirft der Text einen Blick in die Forschungsgeschichte des autonomen Fahrens: Am Beispiel eines experimentellen Sehsystems, das an der Universität der Bundeswehr in München in den 1980er Jahren von Dickmanns entwickelt wurde, wird die Frage beantwortet, ob maschinelle Sehsysteme Blicke werfen und ob sie in die Zukunft schauen können. Außerdem wird drittens untersucht, ob die Bilder im Sensor oder im Rechner entstehen und ob es sich überhaupt um Bilder handelt. Im Ausblick wird dann erörtert, inwiefern die Sehsysteme autonomer Fahrzeuge die Machtverhältnisse zwischen Mensch und Maschine umverteilen – ein Aspekt, der in den meisten Reflektionen zum autonomen Fahren bisher zu wenig diskutiert wird.

4 Virilio, Paul: *The vision machine*, Bloomington/Indianapolis: Indiana University Press 1994, S. 47 (Original: Virilio, Paul: *La machine de vision*, Paris: Éditions Galilée 1988).

Kurze Typologie der Fahrzeugsensorik

In folgendem Abschnitt werden die verschiedenen heute gebräuchlichen Sensoren genauer klassifiziert, kurz historisch eingeordnet und ihre Vor- und Nachteile beschrieben.

Bis zu 100 Sensoren sind heute in modernen Autos verbaut. Erstens gibt es Sensoren, die der *Eigenwahrnehmung* des Fahrzeugs dienen. Sie reichen von Drehzahlsensoren über Temperatur-, Tank- und Ölfüllstandssensoren bis zu Regensensoren. Zweitens gibt es Sensoren, die *Eigen- und Umfeldwahrnehmung* verknüpfen. Mit ihrer Hilfe kann die aktuelle Position – also die eigene Lage relativ zur Umwelt – geschätzt werden: Durch das Messen der Radumdrehungen zeigen odometrische Systeme, wie weit sich ein Fahrzeug fortbewegt hat. Inertial-Sensoren liefern Informationen über die aktuelle Interaktion des Fahrzeugs mit der Umgebung (Beschleunigungs- und Drehrate). Diese Informationen über die Fahrzeugdynamik sind auch für die visuelle Wahrnehmung, die Bilderkennung zwingend notwendig, um die Bilder der Umwelt richtig zu interpretieren.

Im Mittelpunkt der medialen Debatte um das autonome Fahren steht heute jedoch vor allem die dritte Gruppe von Augen, welche die *Fahrzeugumgebung* erfassen – Kameras, Radar, Ultraschall, Infrarot und Lidar. Dem Zeitpunkt ihrer Entwicklung für das Automobil chronologisch folgend sollen hier zunächst die passiven und dann die aktiven Umgebungssensoren vorgestellt werden:

Kameras werden als passive Sensoren bezeichnet, da sie kein eigenes Licht ausstrahlen – sie registrieren Helligkeits- und Farbunterschiede. Mit Hilfe dieser Informationen kann eine Software dann die Größe, Farbe und Kontur von Objekten bestimmen und diese klassifizieren (z.B. Fußgänger oder Mülltonne). Die Idee, ein autonomes Fahrzeug mit Hilfe einer Kamera zu steuern, kam in den 1960er Jahren im Bereich der Robotik auf. So propagierte John McCarthy, einer der Pioniere der Künstlichen Intelligenz und Direktor des Stanford Artificial Intelligence Laboratory den Einsatz von Kameras erstmals 1968 in einem visionären Aufsatz über »Computer Controlled Cars«.⁵ Heute werden Kameras zum Beispiel für das *tracking* anderer Fahrzeuge und die Erkennung von Ampeln eingesetzt. Ein Nachteil von Kameras ist ihre hohe Datenrate. Da sie auf eine gute Beleuchtung angewiesen sind, haben sie Schwierigkeiten mit Nebel, Gegenlicht, Spiegelungen

5 McCarthy, John: »Computer Controlled Cars«, 1968, <http://jmc.stanford.edu/commentary/progress/cars.pdf> vom 24.04.2021.

und Schatten. Ungünstig ist auch, dass einzelne Kameras nur 2D-Ansichten liefern, die von Algorithmen in 3D-Informationen transformiert werden müssen. Stereoskopische Kameras liefern zwei Bilder aus unterschiedlichen Positionen – aber mit exakt gleicher Blickrichtung – die dann in 3-D-Modelle umgewandelt werden.

Im Gegensatz zu passiven Sensoren senden aktive Sensoren Strahlen aus, die von der Umwelt reflektiert werden. Aus der Laufzeit des Signals wird die Entfernung zu einem Objekt ermittelt.

Dieses Prinzip wurde erstmals für Radar (*radio detection and ranging*) entwickelt, das mit elektromagnetischen Wellen im Radiofrequenzbereich arbeitet. Die Geschichte des Radars lässt sich bis zum Ende des 19. Jahrhunderts zurückverfolgen. Erst kurz vor dem 2. Weltkrieg entwickelten mehrere Länder Radar simultan zu militärischen Zwecken. Das neue Ortungsverfahren wurde zur kriegsentscheidenden Technologie der Alliierten im See- und Luftkrieg.⁶ Während der internationale Schiffs- und Flugverkehr schon bald mit Radartechnik und später der Satellitenkommunikation vernetzt wurde, blieb das Automobil lange hinter dieser Entwicklung zurück. Dies liegt vor allem an der Umwelt, in der sich ein Fahrzeug bewegt: sie ist weitaus komplexer und unvorhersehbarer als die Umgebung von Schiffen und Flugzeugen. Die Entwicklung von Radarsystemen für Fahrzeuge begann in den 1960er Jahren. Die ersten Systeme wurden zu Beginn der 1970er Jahre getestet und waren noch in großen kubischen Kästen untergebracht. Erst in den 1990er Jahren wurden mobile Radarsysteme serienreif.⁷

Im Auto bieten Radarsysteme den Vorteil, dass sie im Gegensatz zu Kameras nur geringe Datenraten benötigen, unabhängig vom Wetter funktionieren und sehr gut Abstände zu Objekten außerhalb des Fahrzeugs schätzen können. Short-Range-Radarsysteme werden für die Erfassung der unmittelbaren Fahrzeugumgebung im Umkreis von 30 Metern verwendet (z.B. *blind-spot-warning*). Long-range Radar hat eine Reichweite von bis zu 250 Metern und wird z.B. für Notbremsassistenten verwendet. Zukünftige Systeme werden sich wie ein Auge an den Nah- oder Fernbereich anpassen können.⁸ Die

6 Vgl. Buder, Robert: *The Invention That Changed The World*, New York: Touchstone 1996.

7 Meinel, Holger/Dickmann, Jürgen: »Automotive Radar: From Its Origin to Future Directions«, in: *Microwave Journal* 56/9 (2013), S. 24-40.

8 Fraunhofer-Institut für Organische Elektronik, Elektronenstrahl- und Plasmatechnik, RadarGlass – Vom Autoscheinwerfer zum Radarsensor, Pressemitteilung vom

Erfassung nicht-metallischer Objekte bereitet Radarsystemen aber Schwierigkeiten. Seitliche Mehrwege-Reflektionen führen durch falsche Ortszuordnungen zu häufigen Fehlalarmen. Die geringe Auflösung kann außerdem dazu führen, dass Objekte detektiert, aber nicht klassifiziert werden können.

Die für Einparkhilfen verwendeten Ultraschall-Sensoren, die akustische Frequenzen oberhalb des menschlichen Hörbereichs nutzen, funktionieren nach demselben Prinzip wie Radar: Sie werden von bis zu 10 Meter entfernten Gegenständen reflektiert. Diese auch Sonar genannten Sensoren funktionieren auch nachts oder bei Nebel.

Ein weiterer Sensortyp sind die Infrarotsensoren. Aktive Infrarot-Sensoren erzeugen elektromagnetische Strahlung mit einer niedrigeren Frequenz als sichtbares Licht. Passive Systeme detektieren die Wärmestrahlung von Menschen, Tieren und Objekten mit einer thermografischen Kamera. Beide Systeme können die Fahrzeugumgebung bei Nacht erfassen und werden vor allem zur Fußgängererkennung eingesetzt. Da IR-Sensoren thermische Energie und nicht reflektiertes Licht registrieren, können sie nicht von der Sonne geblendet werden. Außerdem sind sie frei von Interferenzen mit anderen Sensoren. Eine Schwierigkeit ist aber, dass die Wärmestrahlung morgens zu- und abends abnimmt.

Wie Radar wurde auch die Infrarot-Technologie vor allem im militärischen Bereich entwickelt. Im Mittelpunkt stand dabei der Wunsch, Soldaten Nachtsichtfähigkeiten zu verleihen. Seit dem Vietnam-Krieg setzt die US-Airforce Infrarot-Detektoren auch als Zielleitsystem für Raketen ein.⁹ Die ersten Infrarot-Sensoren für Fahrzeuge wurden Anfang der 1970er Jahre getestet, in Deutschland etwa von MAN und Bosch.¹⁰ Ende der 1990er Jahre entstand ein Dual-use-Markt für Infrarot-Sensoren, die militärische Technologie konnte nun auch zivil verwendet werden. Im Jahr 2000 wurde das erste passive Night Vision-System von Raytheon im Cadillac Deville eingeführt.

Lidar (Light Detection and Ranging) ist ein meist Infrarot-Laser basiertes Sensorsystem, das Entfernungen und Geschwindigkeiten optisch messen

29.09.2020, https://www.fep.fraunhofer.de/de/press_media/05_2020.html vom 24.04.2021.

9 Pierotti, Federico/Ronetti, Alessandra: »Beyond human vision, Towards an archaeology of infrared images«, in: NECSUS. European Journal of Media Studies 7/1 (2018), S. 204.

10 Firgau, Walter: »Ohne Fahrer auf richtigem Kurs«, in: Süddeutsche Zeitung vom 06.09.1973.

kann. Wie beim Radar wird dafür die Laufzeit gemessen, die das Signal benötigt, um auf ein Objekt zu treffen und wieder zum Sensor zurückzukehren. Das Entscheidende der Lidar-Technik sind die 360°-Punktwolken, die durch eine Parallelisierung der Laserstrahlen entstehen. Damit können nicht nur die Konturen von Objekten, sondern auch ihre Bewegungen erfasst werden.¹¹ Schon in den 1930er Jahren wurden erste Experimente mit Lichtimpulsen für meteorologische Zwecke durchgeführt. Nach der Erfindung des Lasers 1960 wurde Lidar von der Rüstungsindustrie und der NASA entwickelt. Die Astronauten der Apollo 15-Mission nutzten Lidar 1971, um die Oberfläche des Mondes zu kartieren.¹² Auch in der Archäologie und der Landwirtschaft werden Lidar-Sensoren zur Landvermessung eingesetzt. Heute finden sie sich auch in Tablet-Computern und Staubsaugern. Seit Mitte der 1980er Jahre wurden Lidare für Fahrzeuge getestet, da sie wie Infrarot und Radar zuverlässig bei Dunkelheit funktionieren. Zunächst waren sie wegen ihrer mechanisch rotierenden Spiegelsysteme sehr teuer, neuere Halbleiter-Lidare sind aber preisgünstiger. Im Auto kann Lidar bis zu 250 Meter weit entfernte Objekte in 3D erfassen und hat eine höhere Auflösung als Radarsysteme, was bei hohen Geschwindigkeiten sehr wichtig ist. Da dies in Echtzeit geschieht, kommen Lidarsensoren ohne den Zwischenschritt bildinterpretierender Software aus. Zwar erkennt Lidar im Gegensatz zu Kameras keine Farben, es kann räumliche Objekte und Schatten aber zuverlässiger unterscheiden als Kameras und erfasst Menschen besser als Radar. Deshalb wird es zur Kollisionswarnung eingesetzt. Es kann auch Spurmarkierungen erkennen, indem es die Reflexion der Streifen detektiert. Bei Regen, Nebel und Schnee ist Lidar allerdings störungsanfällig. Zudem ist die laterale Auflösung bei großen Entfernungen gering.

Zusammenfassend kann erstens festgehalten werden, dass sowohl die passiven als auch die aktiven Sensoren Dinge detektieren können, die dem menschlichen Auge verborgen bleiben. Sie erweitern das Spektrum des für den Menschen Sichtbaren. Zweitens ist wichtig, dass der Sehstrahl aller oben erwähnten Sensoren Sichtbarkeit herstellt, aber zugleich unsichtbar

11 Vgl. Sprenger, Florian: Epistemologien des Umgebens, Zur Geschichte, Ökologie und Biopolitik künstlicher Environments, Bielefeld: Transcript 2019, S. 486.

12 Wilcox, Philip: »Vision Correction: Identifying the Best Way for an Autonomous Vehicle to ›See‹ the World«, in: Medium.com vom 09.02.2021, <https://medium.com/s/wlh/vision-correction-identifying-the-best-way-for-an-autonomous-vehicle-to-see-the-world-d633b96b31ac> vom 24.4.2021.

ist. Im Gegensatz zu Auto-Scheinwerfern zeigen Sensoren etwas, ohne selbst gesehen zu werden. Beide Faktoren erklären die starke Implikation militärischer Akteure in der Entwicklungsgeschichte der verschiedenen Sensortechnologien, die hier nur kurz angerissen werden konnte. Die in naher Zukunft zu erwartende Fusion mehrerer Sensorsignale, das Versprechen also, ihre Vorteile zu kombinieren, ist sowohl im automobilen als auch im militärischen Bereich ein logischer Schritt.

Sehmaschinen ohne Blick?

Schauen wir uns nun einen Sensor genauer an – die Kamera. In *Die Sehmaschine* (1988) vertritt Paul Virilio die These, dass computergesteuerte Kameras über ein »Sehen ohne Blick« verfügten.¹³ Demnach besitzen Sehmaschinen also Augen, können aber keine Blicke werfen. Bei diesem blicklosen Sehen handle es sich um eine Form der Erblindung, einen »Nicht-Blick«.¹⁴

Im Folgenden soll diese These am Beispiel eines für das autonome Fahren entwickelten, experimentellen Kamera-Blicksystems überprüft werden. Dafür müssen wir uns zunächst verschiedenen Definitionen des Sehens und des Blicks zuwenden: Sehen bedeutet laut einer Lexikon-Definition, »mit dem Gesichtssinn, mit den Augen optische Eindrücke wahrnehmen«.¹⁵ Unter dem Gesichtssinn wird dabei die »Fähigkeit von Lebewesen« verstanden, »mithilfe bestimmter Organe Lichtsinnesreize aufzunehmen«.¹⁶ Diese enge Definition der visuellen Wahrnehmung ist heute veraltet, da sie zum einen Maschinen die Sehfähigkeit abspricht, zum anderen nicht-optische Sehverfahren wie Radar ausschließt. *Machine Vision* – also ein System, das es Maschinen ermöglicht, Objekte in Bildern zu detektieren und darauf basierend bestimmte Aufgaben durchzuführen¹⁷, ist im Gegensatz zum menschlichen Auge nicht nur auf Licht angewiesen. Es handelt sich hier um eine neue, auf Wissen abzielende Form des Sehens.

13 P. Virilio: The vision machine, S. 59. Virilio hat diese These auch später noch wiederholt: »There is now vision without a gaze« (Armitage, John: Virilio live, Selected Interviews, London/Thousand Oaks/New Delhi: SAGE 2001, S. 41).

14 P. Virilio: The vision machine, S. 73.

15 Duden Wörterbuch, <https://www.duden.de/rechtschreibung/sehen> vom 11.07.2021.

16 Duden Wörterbuch, <https://www.duden.de/rechtschreibung/Gesichtssinn> vom 11.07.2021.

17 Vgl. Myler, Harley R.: Fundamentals of Machine Vision, Bellingham: SPIE 1999, S. 13.

Der Blick geht über die Funktion des Sehens hinaus: Während das Sehen ein eher ungerichtetes, panoramaartiges Herumschauen meint, bezeichnet der Blick ein absichtliches oder aktives, stetiges und zielgerichtetes Schauen. Der Blick kann also als die Art und Weise definiert werden, *wie* Menschen oder Maschinen ein Objekt betrachten. Wichtig ist dabei, dass der Blick die Welt aktiv mit Bedeutung versieht und dabei die Position des Selbst wie des Anderen mitkonstruiert.¹⁸ Der Blick ist also eng mit Fragen von Wissen und Macht verknüpft. Er symbolisiert Christian Kravagna zufolge »eine Herrschaftsinstanz, eine Technik der Distanz und Kontrolle«.¹⁹

Ernst Dieter Dickmanns und der Blick der Maschine

Mit Wahrnehmungsorganen ausgestattete autonome Autos können keine bösen, ängstlichen, abwertenden, neugierigen, verwunderten, dankbaren oder zärtlichen Blicke werfen, ihnen mangelt es an Emotionen. In diesem Sinne – und nur in diesem – sind sie also Sehmaschinen ohne Blick.²⁰ Verstehen wir den Blick hingegen als zielgerichtetes Schauen, können Sehmaschinen durchaus Blicke werfen.

Besonders deutlich wird dies in den Ansätzen des Kognitionswissenschaftlers und Robotikers Ernst Dieter Dickmanns, der sich seit 1977 an der Universität der Bundeswehr in München (UniBwM) mit der visuellen Steuerung von Straßenfahrzeugen befasste. Er erforschte und entwickelte in den 1980er Jahren ein voll autonomes Fahrzeug, das nur durch Auswertung von Kamerabildern dem Straßenverlauf einer Autobahn folgen und auf Hindernisse reagieren konnte. Bei dem Versuchsfahrzeug für autonome Mobilität und Rechnersehen (VaMoRs) handelte es sich um einen Mercedes-Benz 508 D Kastenwagen, der die umfangreiche Computerausrüstung aufnehmen konnte (Abb. 1).

Die meisten internationalen Forschungsgruppen arbeiteten zu dieser Zeit mit einer einzelnen Kamera, die statisch am Fahrzeug fixiert war. Bei hö-

18 Vgl. Sturken, Marita/Cartwright, Lisa: Practices of Looking. An Introduction to Visual Culture, Oxford: Oxford University Press 2001, S. 10.

19 Kravagna, Christian: »Vorwort«, in: Christian Kravagna (Hg.): Privileg Blick, Kritik der visuellen Kultur, Berlin: Edition ID-Archiv 1997, S. 7.

20 Auf bestimmte Sensoren trifft Virilios These durchaus zu: Ein Bewegungsmelder zum Beispiel besitzt keinen Blick, wenn er einfach einen bestimmten Raum erfasst und irgendwo auftretende Helligkeitsunterschiede detektiert.

Abbildung 1: Das Versuchsfahrzeug für autonome Mobilität und Rechnersehen (Va-MoRs), ein 5 Tonnen schwerer Mercedes-Benz 508 D-Kastenwagen, den die Universität der Bundeswehr von 1985-1991 als fahrendes Labor nutzte.



Copyright: Ernst Dieter Dickmanns

heren Geschwindigkeiten müssen weit entfernte Objekte aber frühzeitig erkannt werden und in unübersichtlichen Straßensituationen ist es notwendig, viele sich bewegende Objekte gleichzeitig zu erfassen und zu unterscheiden. Dickmanns Team installierte deshalb zwei Kameras mit verschiedenen Fokusslängen (Tele- und Weitwinkelobjektiv) auf einer Kameraplattform, die eine aktive Blicksteuerung ermöglichte. Die Kamera bewegte sich nicht nur horizontal und vertikal mit dem Fahrzeug sondern konnte sich ihm gegenüber um die Hochachse und die Nickachse drehen (Abb. 2). Dieses Prinzip hatte Dickmanns bereits um 1980 mit einem speziell dafür beschafften Simulationskreis getestet, der bekannten Anlagen für das Pilotentraining nachempfunden war. Das experimentelle Dispositiv machte Medientechniken (Kamera, Leinwand, Projektor) und Reflektionen über die physiologische Funktionsweise des menschlichen Sehapparats für die Automatisierung der Fahraufgabe produktiv.

Abbildung 2: Die Kameraplattform im Versuchsfahrzeug für autonome Mobilität und Rechnersehen (VaMoRs) 1986



Copyright: Ernst Dieter Dickmanns

Dickmanns orientierte sich bei diesem bifokalen, aktiven Sehsystem an den Augen der Wirbeltiere: Wie auch die menschlichen Augen können sie sogenannte Sakkaden ausführen.²¹ Mit diesem Begriff bezeichnete der französische Augenarzt Louis-Émile Javal 1879 schnelle Änderungen der Blickrichtungen beider Augen zwischen verschiedenen Orten im Raum.²² Mit Hilfe der beweglichen Kamera-Plattform sollte dieses Sehverhalten auf technische Systeme übertragen werden. Denn eine Kamera an sich hat keinen Blick, erst durch eine horizontal und vertikal drehbare Vorrichtung wird sie mobil, d.h. aktiv steuerbar und erst durch sie bekommt der Gesichtssinn die Möglichkeit, in verschiedene Blickrichtungen zu schauen.

Möglich wurden damit erstens schnelle Blickwechsel, zweitens langsame Schwenks zur Verfolgung sich bewegender Ziele und drittens eine inertia-le Blickstabilisierung, mit der störende Fahrzeugbewegungen kompensiert werden konnten. Der sakkadische Kamerablick dieses aktiven, dynamischen Blicksystems ermöglichte es, sich bewegende Objekte im zentralen Blickfeld zu halten und weiter entfernte Objekte mit hoher Auflösung zu detektieren.²³

Ab 1996 konnten die Versuchsfahrzeuge der Universität der Bundeswehr auch mit aktiven Kamerabewegungen »über die Schulter blicken« und kreuzende Straßen frühzeitig erfassen, um Abbiegemanöver vorzubereiten.²⁴ Die Kamera-Plattform konnte ein Objekt nun so mit ihrem Blick verfolgen, dass es im Blickfeld blieb. Zudem wurde es möglich, negative Hindernisse wie größere Löcher in unebenem Gelände zu erkennen. Neben der Kameratechnologie brauchte das Forschungsfahrzeug bei diesem System allerdings Hintergrundwissen über seine eigenen Fähigkeiten sowie über die Umwelt, vor allem über die möglichen Erscheinungsformen von Straßen- und Objektmerkmalen. Es kam aber völlig ohne digitale Karten aus. Inertialsensoren lieferten Informationen über Beschleunigungen und Drehgeschwindigkeiten, aus denen kurzfristige zukünftige Ortsbewegungen und Richtungsänderungen des

21 Pellkofer, Martin/Lützel, Michael/Dickmanns, Ernst Dieter: »Vertebrate type Perception and Gaze control for Road Vehicles«, in: Raymond A. Jarvis/Alex Zelinsky (Hg.): *Robotics Research. Springer Tracts in Advanced Robotics 6*, Berlin/Heidelberg: Springer 2003, S. 272.

22 Javal, Louis-Émile: »Essai sur la physiologie de la lecture«, in: *Annales d'Oculistique* 82 (1879), S. 242-253.

23 Dickmanns, Ernst Dieter/Rudolf, Gregor/Lützel, Michael et al.: »EMS-Vision, A Perceptual System for Autonomous Vehicles«, in: *IEEE Transactions on Intelligent Transportation Systems* 3/1 (2002), S. 48.

24 Ebd.

Fahrzeugs berechnet werden konnten. Die Kamerabilder stellten Informationen über die zukünftig zu befahrende Umgebung bereit.

Damit wird erstens deutlich, dass der maschinelle Blick des hier vorgestellten autonomen Fahrzeugs immer aus zwei Elementen besteht – der hardwareseitigen Ausrichtung der Blickrichtung der Kamera und dem softwareseitigen Code, der die Daten auf bestimmte Merkmale hin abtastet. Der Blick ist also nicht nur ein Ergebnis der Kamera-Bewegungen, er findet sich vor allem in der Software, da sie programmiert wurde – oder heute durch *machine learning* trainiert werden kann –, um bestimmte Muster, Strukturen, Linien, Kantenmerkmale und Ecken zu erkennen. Zweitens zeigt sich, dass autonomes Fahren sogar auf mehreren parallelen, synchronisierten Blicken im Plural basiert, denn eine Bilderkennungssoftware kann simultan mehrere Objekte einer Straßensituation detektieren.

Die oben beschriebene aktive Blicksteuerung wurde bisher allerdings nicht in die Serienproduktion übernommen. Heutige Assistenzsysteme verwenden nur Sensoren, die ihre Blickrichtung nicht an die Umgebung anpassen können.²⁵ Gearbeitet wird hingegen an Scheinwerfern mit beweglichen Pupillen, die Fußgängern signalisieren sollen, wohin ein autonomes Fahrzeug blickt.²⁶

Maschinelle Blicke in die Zukunft

Wie die oben angesprochenen Blick-Regime zielen die Blicke des maschinellen Sehens auf Wissen und Macht ab: Sie dienen im hier besprochenen Fallbeispiel dazu, Wissen über die äußere Welt zu generieren und damit die Fahraufgabe zu meistern. Wird dieses Ziel erfolgreich erreicht, handelt es sich zweifellos um ein machtvolleres technologisches System.

Dies zeigt sich besonders in einem weiteren Punkt: Das an der Universität der Bundeswehr in München entwickelte maschinelle Blicksystem ermöglichte nicht nur einen Ausgriff in den Raum, sondern auch in die Zeit. Es ziel-

25 Dickmanns, Ernst Dieter: »BarvEye, Active gaze control for autonomous driving«, in: Proceedings of the 10th International Conference on Computer Vision Theory and Applications, Volume 1 (2015), S. 430.

26 Holley, Peter: »How do you get people to trust autonomous vehicles?«, in: Washington Post vom 29.08.2018, <https://www.washingtonpost.com/technology/2018/08/29/how-do-you-get-people-trust-autonomous-vehicles-this-company-is-giving-them-virtual-eye-s-vom-24.04.2021>.

te nicht nur auf das Erkennen der Gegenwart, sondern auf eine Vorhersage der Zukunft ab. Paul Virilio weist schon 1984 in *Guerre et Cinéma* darauf hin, dass Sensoren nicht nur sichtbar machen, »was hinter dem Horizont ist und was die Nacht verbirgt, sondern vor allem das, was es nicht oder noch nicht gibt«. ²⁷ Für diesen maschinellen Blick, der es ermöglicht, hypothetische Bilder von zeitlich noch nicht eingetretenen Ereignissen zu erzeugen, möchte ich deshalb den Begriff der prädiktiven Wahrnehmung vorschlagen.

Als Ernst Dieter Dickmanns 1977 begann, mit Methoden der Regelungstechnik das Feld der *Machine Vision* zu erforschen, grenzte er sich von Ansätzen der Informatik und der Künstlichen Intelligenz ab, die komplexe räumliche Szenen aus einzeln analysierten Bildern rekonstruierten. Dieses auf eine vollständige Umgebungserfassung setzende Verfahren benötigte aber eine hohe Speicherkapazität und verlängerte die Auswertungszeit. Um dynamische Prozesse zu erfassen, speichert der regelungstechnische Ansatz Dickmanns nur das letzte von der Kamera aufgenommene Bild und konzentriert die Analyse auf relevante Bildausschnitte, zum Beispiel die oben bereits erwähnten Kantenmerkmale von Fahrbahnrandern und Spurmarkierungen. Diese Objekte werden mit geometrischen Modellen klassifiziert. Zugleich wird das Bewegungsverhalten des Fahrzeugs und anderer Objekte mit Differentialgleichungen erfasst, welche die physischen Bewegungsgesetze enthalten. Der Rechner baut aus den Kamerasignalen und seinem Hintergrundwissen dann ein internes Abbild der Außenwelt in Raum und Zeit auf.

Dickmanns Konzept unterscheidet sich damit von anderen Ansätzen, da es neben dem dreidimensionalen Raum auch den Faktor Zeit berücksichtigt – deshalb wird es 4-D-Ansatz genannt. Das aktuell von der Kamera aufgenommene Bild der Umwelt wird mit einer aus der vorherigen Bildfolge abgeleiteten rechnerinternen Vorhersage verglichen. Mit Hilfe bestimmter Modelle für Objekte, Subjekte und Bewegungsprozesse kann dann vorhergesagt werden, wo ein bestimmtes Bildmerkmal im nächsten Videoframe auftauchen könnte. ²⁸ Dies erlaubt es, Vorhersagen über den in Zukunft zu erwartenden Bewegungsverlauf des Fahrzeugs zu treffen. Anstatt nur ein aus der externen

27 Virilio, Paul: Krieg und Kino. Logistik der Wahrnehmung, Frankfurt a.M. : Fischer 1994, S. 176 (Original : Virilio, Paul, *Guerre et cinéma*, Tome 1 : Logistique de la perception, Paris : Éditions Cahiers du cinéma 1984).

28 Dickmanns, Ernst Dieter: »Developing the Sense of Vision for Autonomous Road Vehicles at UniBwM«, in: IEEE computer 50/12 (2017), S. 27.

Welt aufgenommenes Bild der Vergangenheit zu zeigen, erzeugt der Computer also ein rechner-internes Bild der Zukunft. Er kann die Zukunft dabei nicht »sehen«, aber ihre Wahrscheinlichkeit schätzen. Ein Beispiel: Erfasst die Kamera ein Auto, welches das Fahrzeug gerade von links hinten überholt, sagt das Weltmodell voraus, dass dieses Auto in einigen Millisekunden seitlich am Fahrzeug vorbeifahren wird. »Die Information ist nicht mehr starr« wie auf den alten Photographien, können wir mit Paul Virilio hinzufügen.²⁹ Sie erlaubt nun »die Interpretation des Vergangenen und Zukünftigen.«

Bilder von Maschinen für Maschinen

Schließlich stellt sich die Frage, wo diese von dem in die Zeit ausgreifenden maschinellen Blick produzierten Bilder eigentlich entstehen und von wem sie betrachtet werden.

Ein auch für autonome automobile Sehmaschinen bedeutsamer Aspekt von *Machine Vision* ist das Verschwinden des menschlichen Betrachters. Während historische Quellen der visuellen Kultur – anatomische Zeichnungen, mikroskopische oder teleskopische Ansichten, fotografische und filmische Bilder – vor allem für den Blick des menschlichen Auges produziert wurden, werde die überwiegende Mehrheit der Bilder heute »von Maschinen für Maschinen« erzeugt, betonte Virilio in *La machine de vision* schon 1988.³⁰ Der menschliche Beobachter ist aus der direkten oder indirekten Betrachtung dieser synthetischen Bilder herausgenommen,³¹ er ist *out of the loop*, sei eine »Ausnahme von der Regel« geworden, wie der Künstler Trevor Paglen konstatiert.³² Auch Jonathan Crary wies 1992 auf den ersten Seiten seines Buches *Techniken des Betrachters* darauf hin, dass computergenerierte Bilder das Sehen auf einer »vom menschlichen Betrachter getrennten Ebene« neu anordnen³³, was von W.J.T. Mitchell zugespitzt wurde: Diese Technologien

29 P. Virilio: Krieg und Kino, S. 33.

30 P. Virilio: The Vision Machine, S. 60.

31 Ebd.

32 Paglen, Trevor: »Invisible Images (Your Pictures Are Looking at You)«, in: The New Inquiry vom 08.12.2016, <https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you> vom 24.04.2021.

33 Crary, Jonathan: Techniques of the Observer. On Vision and Modernity in the 19th Century, Cambridge/London: MIT Press 1992, S. 1.

würden das Sehen nicht vom »Menschen« abkoppeln, aber »die Bedingungen verändern, unter denen sich das menschliche Sehen artikuliert«.³⁴

Am Beispiel einer Fahrzeugkamera lässt sich diese Rekonfiguration des Sehens sehr gut nachzeichnen: Im Gegensatz zu einer Überwachungskamera, deren in Bilder umgewandelte Signale von menschlichen Angestellten in einem Kontrollzentrum auf einem Monitor betrachtet werden, wird in autonomen Fahrzeugen der Output der Sensoren direkt zum Input eines Rechners: Der Sensor einer Fahrzeugkamera verwandelt Licht in elektrische Ladungen. Diese Messungen (Helligkeits- und Farbwerte) – Echos einer Wechselwirkung von Sensor und Umgebung – werden dann direkt an einen Rechner geschickt, wo sie weiter verarbeitet werden.³⁵ Erst im Rechner werden aus den Daten Bilder, die gespeichert und weitergegeben werden können. Diese Bilder sind also kein Input, sondern ein Output des Rechners. Damit werde die Logik des Kinos umgekehrt, schreibt Simone Arcagni: Anstatt optisch aufgenommene Bilder der Außenwelt zu zeigen, erzeuge maschinelles Sehen mathematisch generierte Wissensbilder der Umwelt. Es handelt sich hier also um ein post-optisches Sehen, das keine Abbildungen der Welt produziert, sondern Visualisierungen von Wissen.

»Jedes Bild verkörpert eine bestimmte Art des Sehens.«³⁶ Diese These belegte John Berger 1972 mit einem Verweis auf die Fotografie – in der Wahl des Motivs, des Bildausschnitts, der Belichtung etc. zeigen sich die Sehinteressen des Fotografen. Maschinell erzeugte Bilder verkörpern eine ganz andere Form des Sehens: Diese *operativen Bilder*, wie sie der Filmemacher Harun Farocki 2001 in seiner Werkreihe *Auge/Maschine* nannte,³⁷ repräsentieren keine Objekte, sondern sind Teil technischer Operationen. Sie zeigen nicht die Welt, sondern »die apparative Vorstellung derselben« schreibt Wolfgang Ernst.³⁸ In ihnen zeigen sich die Blickinteressen der *Machine Vision*.

Die erzeugten Bilder liegen in einer diskreten Form vor und unterscheiden sich damit von analogen Fotografien ebenso grundlegend wie Fotografien

34 Mitchell, W.J.T.: »Der Pictorial Turn«, in: Christian Kravagna (Hg.): *Privileg Blick, Kritik der visuellen Kultur*, Berlin: Edition ID-Archiv 1997, S. 26.

35 Bei manchen Sensoren (CMOS) findet die Signalentrauschung und -verarbeitung allerede schon im Sensor statt.

36 Berger, John: *Sehen, Das Bild der Welt in der Bilderwelt*, Reinbek bei Hamburg: Rowohlt 1974, S. 10.

37 Farocki, Harun: »Phantom Images«, in: *Public 29* (2004), S. 17.

38 Ernst, Wolfgang: »Der medienarchäologische Blick«, in: Harro Segeberg (Hg.): *Die Medien und ihre Technik. Theorien – Modelle – Geschichte*, Marburg: Schüren 2004, S. 29.

von Malereien, beobachtete der früher am MIT forschende Architekt William J. Mitchell in seinem Buch *The Reconfigured Eye*.³⁹ Verstehen wir Bilder in einer sehr engen Definition als flächige, figurative Visualisierungen, sind digitale Bilder also gar keine Bilder, da sie algorithmisch erzeugt und in Form eines binären Codes gespeichert sind. Bei diesen von informationsgebenden Verfahren erzeugten Daten handelt es sich um unsichtbare, wir könnten auch sagen potentielle Bilder. Nur wenn sie auf einem Monitor dargestellt werden, bekommen sie eine bildliche Erscheinung – dann aber wiederum in Form analoger Bilder, die digital erzeugte Daten visuell darstellen.⁴⁰

Im Gegensatz zur Fotografie kommen maschinelle Sehsysteme also ohne humanen Akteur sowohl auf der Ebene des Senders, als auch auf der Ebene des Empfängers aus; sie sind vollständig selbstreferenziell.

Ausblick

Zum Schluss sei darauf hingewiesen, dass dieser *maschinelle Blick*, diese in die Zeit ausgreifenden, die Zukunft bewirtschaftenden *Wissens- oder Datenbilder*, diese Bildproduktionen von *Maschinen für Maschinen* in autonomen Fahrzeugen dazu dienen, von Algorithmen und probabilistischen Modellen gesteuerte Entscheidungen zu treffen.

Diese Entwicklung erzeugt vor allem den Eindruck, dass sich uns hier etwas zu entziehen beginnt. Wir werden Zeugen einer zunehmenden Eigenmacht von Technologien, »die die Kapazitäten des Menschen unterlaufen und ihn als entscheidende Instanz in Frage stellen« wie Florian Sprenger in der Einleitung zu diesem Band bemerkt hat. Kurz gesagt zeigt sich in den Blicksystemen autonomer Fahrzeuge also eine Neuverteilung von Handlungsmacht.

Lucy Suchman und Jutta Weber vertreten die These, dass Handlungsmacht nicht als isolierte Fähigkeit eines singulären Akteurs gedacht werden

39 Mitchell, William J.: *The Reconfigured Eye, Visual Truth in the Post-Photographic Era*, Cambridge/London: MIT Press 1992, S. 4.

40 Vgl. Pias, Claus: »Das digitale Bild gibt es nicht. Über das (Nicht-)Wissen der Bilder und die informatische Illusion«, in: *Zeitenblicke* 2/1 (2003), <https://www.zeitenblicke.historicum.net/2003/01/pias/index.html> vom 24.04.2021.

kann.⁴¹ *Agency* entstehe immer in einem relationalen, kontingenten Netzwerk aus humanen und nicht-humanen Akteurskonfigurationen. Auch Florian Sprenger weist darauf hin, dass Handlungsmacht sich nicht an einem Ort lokalisieren lasse. In autonomen Fahrzeugen entstehe sie in einer verteilten kognitiven *assemblage* aus Sensordaten, Filteralgorithmen, den daraus abgeleiteten Wahrscheinlichkeiten, menschlichen Akteuren und der jeweiligen Umgebung. Das Spektrum der an diesen »Mikroentscheidungen« beteiligten Akteure werde damit stark ausgeweitet und gehe über fest determinierte Algorithmen hinaus.⁴²

Diese vor allem von den Science und Technology Studies hervorgebrachte Perspektive bietet einerseits einen produktiven Ausweg aus der dichotomischen Trennung von Mensch und Maschine, wie sie bei Virilio sehr stark ausgeprägt ist. Die Metapher einer *assemblage*, eines *ensembles* aus Mensch und Maschine birgt andererseits aber das Risiko, die fundamentalen Verschiebungen zu vernachlässigen, die sich in der Entwicklung autonomer Sehsysteme verbergen. Es sind Zweifel angebracht, ob diese Perspektive wirklich den *shift*, den Bruch zwischen einem manuell gesteuerten Automobil und einem vollautonomen Fahrzeug erfasst, das sogar ohne Insassen seinen Weg finden soll.

Die zentrale Frage lautet, welche Transformationen die *humanen Akteure* in dieser *assemblage* durchlaufen. An welcher Stelle und zu welchem Zeitpunkt werden sie in autonom entscheidenden Sehmaschinen überhaupt noch an Entscheidungen mitwirken? Eine Teilantwort lässt sich heute schon geben: Die Handlungsmacht der Fahrerin oder des Fahrers verlagert sich nicht nur in den Rechner, sondern auch auf eine Kontrolltätigkeit von Dritten, die das Fahrzeug ständig überwachen. So werden die seit Oktober 2020 ohne Sicherheitsfahrer:in agierenden Versuchsfahrzeuge von Waymo in Phoenix permanent von menschlichem Kontrollpersonal des Unternehmens überwacht. Die zunehmend autonom werdenden Seh- und Fahrmaschinen bleiben gleichzeitig also auf menschliche Arbeit angewiesen. Sam Hind hat dies in diesem Band am Beispiel der *remote workers* oder *ghost operators* gezeigt. Jan Distelmeyer hat auf die unbezahlte Datenarbeit von Tesla-Fahrer:innen hingewie-

41 Suchman, Lucy/Weber, Jutta: »Human-Machine Autonomies«, in: Nehal Bhuta/Susanne Beck/Robin Geiß et al. (Hg.): *Autonomous Weapons Systems, Law, Ethics, Policy*, Cambridge: Cambridge University Press 2016, S. 20f.

42 Sprenger, Florian: »Microdecisions and autonomy in self-driving cars: virtual probabilities«, in: *AI & Society* 7/5 (2020), S. 176-190.

sen. Auch bei Fahrzeugen, die *machine vision* für autonome Entscheidungen nutzen, bleibt der Mensch also – vorerst – *in the loop*.

Sichtbar wird aber schon heute, dass die Distanz zwischen humanen und non-humanen Agenten zunimmt. Sie erhalten in dieser neuen Form von *assemblage* nicht die gleichen Mitspracherechte. Sonst hätten autonome Fahrzeuge auch keinen Sinn – ihr Zweck ist es ja gerade, den Menschen aus dem Prozess der Steuerung zu entfernen, vor allem, um Unfälle zu verhindern. Die Entscheidungen dieser sensorisch aufgerüsteten autonomen Systeme gehen dabei aber über Beschleunigungs-, Lenk-, und Bremsmanöver hinaus – sie entscheiden, »wer mit wem kommunizieren darf und wer nicht, wer sich wohin bewegen darf und wer nicht, wer einen Unfall überlebt und wer nicht«. ⁴³ Damit seien sie hochgradig politische Maschinen, die »eine neue Form von Macht« verkörperten, gibt Sprenger zu bedenken. Und diese Macht ist ungleich verteilt, ließe sich hier noch präzisieren: der Anteil der non-humanen Akteure im *decision making* weitet sich immer mehr aus.

Da diese maschinell erzeugten Entscheidungen zunehmend intransparent werden, fordern Anthony McCosker und Rowan Wilken in ihrem Buch *Automating Vision*, an einer neuen *literacy*, einem tiefergehenden Verständnis von *machine vision* zu arbeiten. ⁴⁴ Diese *literacy* dürfe sich nicht nur auf die technische Funktionsweise dieser Bildmaschinen beschränken, sondern müsse die »Kontexte, Konsequenzen, Beziehungen zwischen Systemen, ihre Begrenzungen und möglichen Gebräuche« begreifen. ⁴⁵ Eine kritische Medien- und Kulturanalyse sollte sich deshalb in Zukunft zum Beispiel verstärkt der Frage widmen, welche Auswirkungen autonom entscheidende Sehmaschinen in anderen, über den Autoverkehr hinausgehenden Anwendungsbereichen haben werden.

43 Ebd.

44 McCosker, Anthony/Wilken, Rowan: *Automating Vision. The Social Impact of the New Camera Consciousness*, New York/London: Routledge 2020, S. 5.

45 A. McCosker und R. Wilken: *Automating Vision*, S. 7.

