

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro

(43) Internationales Veröffentlichungsdatum
23. Juni 2022 (23.06.2022)



(10) Internationale Veröffentlichungsnummer
WO 2022/128014 A1

(51) Internationale Patentklassifikation:
G06V 20/56 (2022.01) G06V 10/82 (2022.01)

(21) Internationales Aktenzeichen: PCT/DE2021/200236

(22) Internationales Anmeldedatum:
03. Dezember 2021 (03.12.2021)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
10 2020 215 860.6
15. Dezember 2020 (15.12.2020) DE

(71) Anmelder: CONTI TEMIC MICROELECTRONIC
GMBH [DE/DE]; Ringerlerstraße 17, 85057 Ingolstadt (DE).

(72) Erfinder: SCHARFENBERGER, Christian; c/o Conti
Temic microelectronic GmbH, Intellectual Property, Nord-
ostpark 30, 90411 Nürnberg (DE). KARG, Michelle; c/o
Conti Temic microelectronic GmbH, Intellectual Property,
Nordostpark 30, 90411 Nürnberg (DE).

(81) Bestimmungsstaaten (soweit nicht anders angegeben, für
jede verfügbare nationale Schutzrechtsart): AE, AG, AL,
AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN,
KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO,
NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,
SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(54) Title: CORRECTION OF IMAGES FROM A PANORAMIC-VIEW CAMERA SYSTEM IN THE CASE OF RAIN, INCIDENT
LIGHT AND CONTAMINATION

(54) Bezeichnung: KORREKTUR VON BILDERN EINES RUNDUMSICHTKAMERASYSTEMS BEI REGEN, LICHTEINFALL
UND VERSCHMUTZUNG

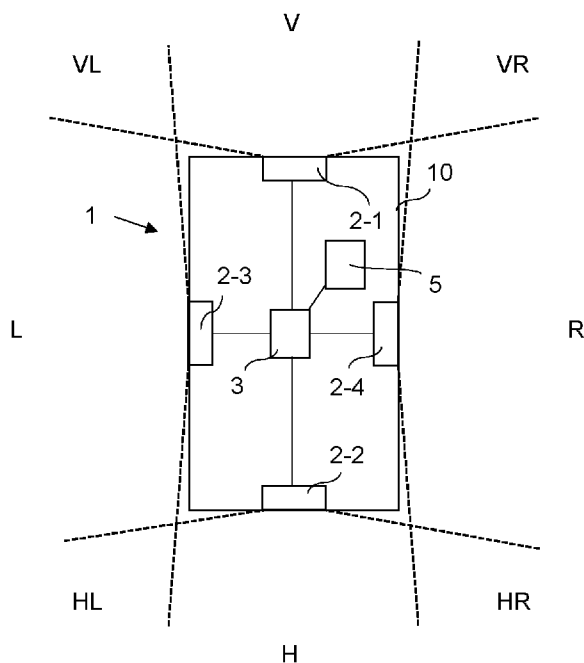


Fig. 2

(57) Abstract: The invention relates to a machine learning method, a method and a device for correcting input image data (Ini) from a plurality of cameras (2-i) of a panoramic-view system, for example a vehicle-borne environment-detection camera system, which data are impaired by rain, incident light and/or dirt. The method for correcting input image data from a plurality of cameras (2-i) of a panoramic-view system, which data are impaired by rain, incident light and/or dirt, comprises the steps of: a) input image data (Ini) detected by the cameras (2-i), which data are impaired by rain, incident light and/or dirt, are provided to a trained artificial neural network (CNN1, CNN10, CNN11, CNN12), b) the trained artificial neural network (CNN1, CNN10, CNN11, CNN12) is configured to convert the input image data (Ini) with impairment due to rain, incident light, and/or dirt into corrected output image data (Opti) without impairment, and to determine a degree of certainty c which characterises, according to the degree of wetting by water, incident light, and/or contamination for an image of the input image data, the certainty of the network that the image correction of the network is accurate, and c) the trained artificial neural network (CNN1, CNN10, CNN11, CNN12) is configured to output the output image data (Opti) and the determined degree of certainty c. The method advantageously allows object identification when cameras are fogged up and generation of an image data stream for human and computer vision from a network for an optimised correspondence search.

(57) Zusammenfassung: Die Erfindung bezieht sich auf ein maschinelles Lernverfahren, ein Verfahren und eine Vorrichtung zum Korrigieren von Eingangs-Bilddaten (Ini), die durch Regen,



(84) **Bestimmungsstaaten** (soweit nicht anders angegeben, für jede verfügbare regionale Schutzrechtsart): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), eurasisches (AM, AZ, BY, KG, KZ, RU, TJ, TM), europäisches (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

- mit internationalem Recherchenbericht (Artikel 21 Absatz 3)
- vor Ablauf der für Änderungen der Ansprüche geltenden Frist; Veröffentlichung wird wiederholt, falls Änderungen eingehen (Regel 48 Absatz 2 Buchstabe h)

Lichteinfall und/oder Schmutz beeinträchtigt sind, einer Mehrzahl von Kameras (2-i) eines Rundumsichtsystems, beispielsweise eines fahrzeuggebundenen Umfelderkassungskamerasystems. Das Verfahren zum Korrigieren von Eingangs-Bilddaten einer Mehrzahl von Kameras (2-i) eines Rundumsichtsystems, die durch Regen, Lichteinfall und/oder Schmutz beeinträchtigt sind, umfasst die Schritte: a) von den Kameras (2-i) erfasste Eingangs-Bilddaten (Ini), die durch Regen, Lichteinfall und/oder Schmutz beeinträchtigt sind, werden einem trainierten künstlichen neuronalen Netzwerk (CNN1, CNN10, CNN11, CNN12) bereitgestellt, b) das trainierte künstliche neuronale Netzwerk (CNN1, CNN10, CNN11, CNN12) ist dazu konfiguriert, die Eingangs-Bilddaten (Ini) mit Regen-, Lichteinfall und/oder Schmutzbeeinträchtigung in korrigierte Ausgabe-Bilddaten (Opti) ohne Beeinträchtigung umzuwandeln und ein Sicherheitsmaß c zu ermitteln, das abhängig vom Grad der Benetzung durch Wasser, Lichteinfall und/oder Verschmutzung für ein Bild der Eingangs-Bilddaten die Sicherheit des Netzwerks charakterisiert, dass die Bildkorrektur des Netzwerks zutreffend ist und c) das trainierte künstliche neuronale Netzwerk (CNN1, CNN10, CNN11, CNN12) ist dazu konfiguriert, die Ausgabe-Bilddaten (Opti) und das ermittelte Sicherheitsmaß c auszugeben. Vorteilhaft ermöglicht das Verfahren eine Objekterkennung bei Beschlag von Kameras und eine Generierung eines Bilddatenstromes für Human und Computer Vision aus einem Netzwerk für eine optimierte Korrespondenzsuche.

Korrektur von Bildern eines Rundumsichtkamarasystems bei Regen, Lichteinfall und
Verschmutzung

Die Erfindung bezieht sich auf ein maschinelles Lernverfahren, ein Verfahren und eine
5 Vorrichtung zur Korrektur von Bilddaten einer Mehrzahl von Fahrzeugkamas eines
Rundumsichtsystems bei Regen, Lichteinfall oder Verschmutzung, beispielsweise
eines fahrzeuggebundenen Rundumsichtkamarasystems.

Heutige und zukünftige Fahrzeuge sind mit kamerabasierten Assistenzsystemen
10 ausgestattet, welche der Erkennung von Objekten zur Vermeidung von Kollisionen und
der Erkennung von Straßenbegrenzungen zum Halten des Fahrzeuges in der Spur
dienen. Dazu werden beispielsweise nach vorne schauende Kameras verwendet.
Neben nach vorne schauenden Kameras werden ebenfalls Surroundview-
(Rundumsicht-) oder Satellitenkamas verwendet, die in beliebiger Anordnung an
15 einem Fahrzeug Detektionsfunktionen für Fahren, Parken oder für eine Visualisierung
im nahen und fernen 360° Umfeld (oder Teilen davon) um das Fahrzeug realisieren.

Es gibt Konzepte für nach vorne schauende Kameras, die eine Detektionsfunktion zur
20 Detektion von Regen umfassen. Kombiniert mit einer kamerabasierten
Taglichterkennung zur Steuerung der Fahrzeugscheinwerfer kann mit einer Kamera ein
sogenannter Rain Light Detektor realisiert werden, der Regen auf der
Windschutzscheibe erkennt und beispielsweise den Scheibenwischer aktiviert.

Die auf diesen Kamerasystemen basierenden Erkennungsalgorithmen kombinieren
25 teilweise bereits Ansätze der klassischen Bildverarbeitung mit Ansätzen aus dem
maschinellen Lernen, insbesondere dem Deep Learning. Klassische Ansätze zur
Erkennung von Objekten oder Strukturen im Rahmen der Bildverarbeitung basieren auf
manuell ausgewählten Features (Merkmale), während auf Deep Learning basierende
Ansätze relevante Features im Trainingsverfahren selbst bestimmen und optimieren.

30 Die oben genannten Kamerasysteme zeigen eine Degradation sowohl in der
Erkennung von Objekten als auch bei der Darstellung der Umgebung oder von
Objekten tritt jedoch ein, sobald in einer Situation die Sicht durch Regen, Lichteinfall
oder Schmutz beeinträchtigt ist.

Ist die Sicht der beispielsweise im Spiegelfuß des Innenspiegels eingebauten Frontkamera durch Wassertropfen auf der Scheibe oder Schmutz eingeschränkt, kann die Sicht durch Betätigung des Scheibenwischers wiederhergestellt werden. Dies setzt eine Einbauposition der Kamera im Wischbereich des Scheibenwischers voraus.

5

Durch den steigenden Automatisierungsgrad von Fahrzeugen und Fahrfunktionen kann eine Erhöhung der Anzahl der nach vorne schauenden Kameras vorteilhaft sein. Diese können nicht nur zentriert im Spiegelfuß angebracht sein, sondern als Satellitenkameras nahe der A-Säule in den oberen Ecken einer Windschutzscheibe. Diese Bereiche sind für Detektionsfunktionalitäten eher kritisch, da sie sich außerhalb des Wischbereiches des Scheibenwischers befinden. Eine durch Regentropfen oder Schmutz eingeschränkte Sicht wirkt sich nachteilig auf Detektionsfunktionen dieser Satellitenkameras aus.

10

15

Bei einem steigenden Grad der Automatisierung von Fahrzeugen bis SAE Level 4 / Level 5 werden zunehmend an den Seiten eines Fahrzeuges angebrachte Kameras verwendet, die neben einer Anzeige der Umgebung auch der Detektion von seitlichen Objekten dienen. Häufig werden diese Kameras an der Außenseite des Fahrzeugs angebracht, z.B. im Bereich der Außenspiegel. Sind die (äußeren) Linsen der Kameras durch Wassertropfen oder Schmutz benetzt, kann auch hier die Anzeige- oder Detektionsfunktionalität sehr eingeschränkt sein. Aufgrund von fehlenden Reinigungsmöglichkeiten wie Scheibenwischern führt dies zu einer Degradation oder zu einem Ausfall eines Systems.

20

25

Als letztes Beispiel seien Rückfahrkameras genannt, die meist über dem Nummernschild verbaut sind, und sehr schnell verschmutzen. Auch hier kann es bei Regen oder Staub zu Beschlag kommen, welcher eine saubere Anzeige erschwert.

30

Während CNN-basierte Verfahren zur Objekterkennung weitestgehend in der Lage sind, Verschmutzung oder Benetzung der Linsen durch Wassertropfen zumindest bis zu einem gewissen Grad zu kompensieren, erleben Verfahren zur Objekterkennung, die auf Bildmerkmalen (Features) basieren, wie z.B. optischer Fluss oder Structure-From-Motion, durch Verschmutzung eine starke Degradation.

Bekannt sind algorithmische Verfahren zur Detektion von Schmutz oder Niederschlag auf der äußeren Linse einer Kamera oder auf der Windschutzscheibe eines Fahrzeugs mittels Bildverarbeitung.

5 WO 2013/083120 A1 zeigt ein Verfahren zur Auswertung von Bilddaten einer Fahrzeugkamera, bei dem Informationen über Regentropfen auf einer Scheibe im Sichtbereich der Fahrzeugkamera bei der Auswertung der Bilddaten berücksichtigt werden. Die Informationen über Regentropfen können ihrerseits aus den Bilddaten ermittelt werden. Als Beispiel für die Auswertung der Bilddaten wird eine Erkennung von
10 Objekten genannt, die dann die Informationen gezielt berücksichtigt. Aus einer erkannten Regenstärke kann beispielsweise die Beeinflussung der von der Kamera gesehenen Kanten (Hell/Dunkel oder Farbübergänge) abgeschätzt werden. Entsprechend können kantenbasierte Auswertungserfahren in ihren Schwellwerten angepasst werden. Insbesondere kann aus den Informationen ein Gütekriterium der
15 Bilddaten abgeleitet werden, das dann bei der Auswertung der Bilddaten berücksichtigt wird.

H. Porav et al. zeigen in „I Can See Clearly Now: Image Restoration via De-Raining“, 2019 IEEE Int. Conference on Robotics and Automation (ICRA), Montreal, Canada, p.
20 7087-7093, abgerufen am 13.7.2020 unter:
http://www.robots.ox.ac.uk/~mobile/Papers/ICRA19_porav.pdf ein Verfahren zur Verbesserung von Segmentierungsaufgaben auf Bildern, die durch aufliegende Regentropfen oder Schlieren beeinträchtigt sind. Dazu wurde ein Stereo-Datensatz generiert, bei dem eine Linse durch reale Wassertropfen beeinträchtigt war und die
25 andere Linse klar von Beeinträchtigungen war. Damit wurde ein „Denoising-Generator“ trainiert, um den Effekt der Wassertropfen im Kontext der Bildrekonstruktion und Fahrbahnmarkierungs-Segmentierung zu entfernen.

Ein System wäre wünschenswert, welches algorithmisch eine Aufwertung der Bilder
30 trotz Verschmutzung, Lichteinfall oder Wassertropfen bietet, beispielsweise um eine nachgelagerte Objekterkennung zu verbessern, und zusätzlich eine Funktion zur Regen- und Lichterkennung (Rain + Light Detection) ermöglicht.

Es ist Aufgabe der vorliegenden Erfindung, Lösungen hierfür bereitzustellen.

35

Die Aufgabe wird gelöst durch die Gegenstände der unabhängigen Patentansprüche. Vorteilhafte Ausführungsformen sind Gegenstand der abhängigen Ansprüche, der folgenden Beschreibung sowie der Figuren.

- 5 Ein erfindungsgemäßes Verfahren zum maschinellen Lernen betrifft eine Bildkorrektur von Eingangs-Bilddaten einer Mehrzahl von Kameras eines Rundumsichtsystems, die durch Regen, Lichteinfall und/oder Schmutz beeinträchtigt sind, in korrigierte Ausgabe-Bilddaten mittels eines künstlichen neuronalen Netzwerks. Das Lernen erfolgt mit einer Vielzahl von Trainings-Bildpaaren derart, dass am Eingang des künstlichen
- 10 neuronalen Netzwerks jeweils ein erstes Bild (bzw. erste simultan erfasste Bilder) mit Regen-, Lichteinfall und/oder Schmutzbeeinträchtigung und als Soll-Ausgabe-Bild ein zweites Bild (bzw. zweite simultan zu erzielende Soll-Bilder) derselben Szene ohne Beeinträchtigung bereitgestellt wird. Das künstliche neuronale Netzwerk ist derart ausgelegt, dass es ein Sicherheitsmaß c ermittelt, das vom Grad der Benetzung durch
- 15 Wasser, Lichteinfall und/oder Verschmutzung für ein Eingangs-Bild abhängt. Die Auslegung des Netzwerks kann beispielsweise durch ein entsprechendes Design bzw. eine entsprechende Architektur des künstlichen neuronalen Netzwerks erfolgen. Nach Abschluss des maschinellen Lernens kann das künstliche neuronale Netzwerk das Sicherheitsmaß c für ein neues Eingangsbild (bzw. für jedes von simultan erfassten
- 20 Eingangsbilder der mehreren Kameras) bestimmen und ausgeben. Das Sicherheitsmaß c hängt also vom Grad der Beeinträchtigung durch eine Benetzung mit Regen bzw. Wasser, durch Lichteinfall und/oder durch Verschmutzung ab und charakterisiert beim Einsatz des trainierten Netzwerks die Sicherheit, dass eine Bildkorrektur zutreffend ist.
- 25 Mit anderen Worten charakterisiert das Sicherheitsmaß c gewissermaßen die „(Un-)Sicherheit“, mit der eine Bildkorrektur durch das trainierte neuronale Netzwerk erfolgt. Das Sicherheitsmaß c ist praktisch ein Maß für die Sicherheit des Netzwerks in seine berechnete Ausgabe (d.h. die durch das Netzwerk vorgenommene Bildkorrektur).
- 30 Das künstliche neuronale Netzwerk kann beispielsweise ein Convolutional Neural Network („faltendes neuronales Netzwerk“, CNN) sein.

Die Umwandlung in Ausgabe-Bilddaten ohne Beeinträchtigung schließt i.d.R. die Umwandlung in Ausgabe-Bilddaten mit verringerter Beeinträchtigung mit ein.

Die Kamera kann beispielsweise eine in oder an einem Fahrzeug befestigte (monokulare) Kamera sein, die das Umfeld des Fahrzeugs erfasst. Ein Beispiel einer solchen fahrzeuggebundenen Kamera ist eine hinter der Windschutzscheibe im Inneren des Fahrzeugs angeordnete Kamera, die den vor dem Fahrzeug liegenden Bereich der Fahrzeugumgebung durch die Windschutzscheibe erfassen und abbilden kann.

Die Auswirkung einer Beeinträchtigung eines Kamerabilds durch Regen, Lichteinfall oder Schmutz ist insofern ähnlich, als sie zu einer (lokalen) Unschärfe im Bild führt. In all diesen Fällen ist eine Bildkorrektur erstrebenswert, die diese Unschärfe im Bild verringert bzw. eliminiert.

Gemäß einer Ausführungsform wird mindestens ein Faktor d als Maß für den Unterschied zwischen dem korrigierten Ausgangsbild und dem beeinträchtigten Eingangsbild ermittelt und dem künstlichen neuronalen Netzwerk im Rahmen des Trainings bereitgestellt. Der Faktor d wird von dem künstlichen neuronalen Netzwerk beim Lernen berücksichtigt, beispielsweise in der Weise, dass das neuronale Netzwerk die Verknüpfung von Eingangsbild, Ausgabebild und Faktor d trainiert. Dadurch kann das trainierte Netzwerk später für ein aktuell erfasstes beeinträchtigtes Kamerabild einen Faktor d schätzen oder bestimmen und ein entsprechend stark korrigiertes Ausgabebild erzeugen (bzw. rekonstruieren). Nach Abschluss des Trainings kann also beispielsweise dem trainierten neuronalen Netzwerk ein Faktor d vorgegeben werden und dadurch der Grad der Korrektur des aktuell erfassten Kamerabildes gesteuert werden.

Der Faktor d kann beispielsweise mittels eines lokalen Vergleichs eines ungestörten Bildes mit dem eines durch Regen oder Schmutz beeinträchtigten Bildes ermittelt werden.

Dabei kann man mit Hilfe von 2D-Filtern, die beispielsweise in den Eingangslayern eines künstlichen neuronalen Netzwerks abgebildet sein können, den Faktor d bestimmen.

Im einem einfachen Ausführungsbeispiel lässt sich der Faktor d als Varianz eines 2D Tiefpassfilters darstellen. Alternativ sind komplexere Kontrastwerte (Structural Similarity) oder Korrelationen (Sum of absolute distances – SAD, Sum of squared distances – SSD, Zero-means normalized cross correlation – ZNCC), die aus den beiden Bildern mit lokalen Filtern berechnet werden, ebenso vorstellbar.

Beispielsweise entsteht bei einer durch Regen oder Schmutz verunreinigten Linse einer Kamera ein gestörtes Bild, welches eine Objekterkennung möglicherweise erschwert. Im Rahmen des maschinellen Lernverfahrens kann ein Faktor d aus einem Vergleich von Soll-Ausgabe-Bild und zugehörigem beeinträchtigtem Eingangsbild bestimmt werden. Diese Bestimmung kann vorab erfolgen, d.h. für jedes Trainings-Bildpaar liegt bereits ein Faktor d vor. Alternativ kann die Bestimmung des Faktors d rein auf Grundlage der Trainings-Bildpaare im Rahmen des Lernverfahrens erfolgen.

Durch den Faktor d kann ein Wert bereitgestellt werden, der den Grad einer möglichen Rekonstruktion des korrigierten Bildes anzeigt und anschließenden Bildverarbeitungs- oder Bildanzeigefunktionen mitgegeben wird. Ein niedriger Wert kann beispielsweise eine hohe Korrektur, ein hoher Wert eine niedrige Korrektur für die weiteren Verarbeitungsstufen anzeigen und bei der Ermittlung der Qualität der erzeugten Objektdaten – ebenso wie das Sicherheitsmaß c – berücksichtigt werden.

In einer Ausführungsform werden die Trainings-Bildpaare erzeugt, indem jeweils ein erstes Bild mit Regen-, Lichteinfall und/oder Schmutzbeeinträchtigung (im optischen Pfad der Kamera) und ein zweites Bild ohne Beeinträchtigung zeitgleich oder unmittelbar aufeinander folgend mit unterschiedlichen Belichtungszeiten simultan mit den Kameras aufgenommen bzw. von den Kameras erfasst werden.

In einer Ausführungsform wird (nur) ein künstliches neuronales Netzwerk für alle Fahrzeugkameras gemeinsam bzw. gleichzeitig trainiert.

Es kann jeweils eine Folge von aufeinanderfolgenden Bildern für jede Einzelkamera für das gemeinsame Training verwendet werden.

Die zeitliche Korrelation von Bildern kann beim Training und/oder beim Einsatz des trainierten Netzwerks gewinnbringend berücksichtigt werden.

Informationen über Bildmerkmale und deren Soll-Ausgabe-Bilddaten können verwendet werden, die zu einem Zeitpunkt t von einer Frontkamera und zu einem späteren Zeitpunkt von einer seitlichen Kamera oder der Rückkamera erfasst werden. Dadurch kann trainiert werden, dass ein Objekt mit bestimmten Bildmerkmalen in den Ausgabe-Bildern aller Einzelkameras eine identische Helligkeit und Farbigkeit aufweist.

Gemäß einer Ausführungsform enthalten die Trainings-Bildpaare mindestens eine Folge von aufeinander folgenden Eingangs- und Ausgangsbildern (als Bilddaten). Mit

anderen Worten werden als Bilddaten Bildfolgen (Videosequenzen) eingesetzt. Für das maschinelle Lernen ist in diesem Fall mindestens eine Eingangs-Videosequenz und eine Soll-Videosequenz erforderlich.

Bei der Verwendung von Bildfolgen können vorteilhaft zeitliche Aspekte bzw. Zusammenhänge in der Rekonstruktion (bzw. Bildkorrektur) berücksichtigt werden. Als Beispiel seien Regentropfen oder Schmutzpartikel genannt, die sich mit der Zeit bewegen. Dadurch entstehen Bereiche im Bild, die zu einem Zeitpunkt t eine klare Sicht und zu einem Zeitpunkt $t+1$ eine durch Regen gestörte Sicht enthielten. Durch die Verwendung von Bildfolgen können Informationen in den klaren Bildbereichen zur Rekonstruktion in den durch Regen oder Schmutz gestörten Sichtbereichen verwendet werden.

Der zeitliche Aspekt kann insbesondere bei den durch Schmutz bedeckten Bereichen helfen, ein klares Bild zu rekonstruieren. In einem Beispiel seien Bereich der Linse durch Schmutz bedeckt, und andere Bereiche frei. Zu einem Zeitpunkt t kann ein Objekt vollständig gesehen werden, zu einem anderen Zeitpunkt $t+1$ verhindert Schmutz die vollständige Aufnahme des Objektes. Durch Bewegung des Objektes und/oder der Bewegung der Kamera beim Fahren kann nun die gewonnene Information über das Objekt im Bild zum Zeitpunkt t helfen, das Bild zu einem Zeitpunkt $t+1$ zu rekonstruieren.

In einer Ausführungsform weist das künstliche neuronale Netz eine gemeinsame Eingangsschnittstelle für zwei separate Ausgangsschnittstellen auf. Die gemeinsame Eingangsschnittstelle weist geteilte Merkmalsrepräsentationsschichten auf. An der ersten Ausgangsschnittstelle werden korrigierte (d.h. umgewandelte) Bilddaten ausgegeben. An der zweiten Ausgangsschnittstelle werden ADAS-relevante Detektionen mindestens einer ADAS-Detektionsfunktion ausgegeben. ADAS steht für fortschrittliche Systeme zum assistierten oder automatisierten Fahren (Englisch: Advanced Driver Assistance Systems). ADAS-relevante Detektionen sind somit z.B. Objekte, Gegenstände, Verkehrsteilnehmer, die für ADAS/AD Systeme wichtige Eingangsgrößen darstellen. Das künstliche neuronale Netzwerk umfasst ADAS-Detektionsfunktionen, z.B. eine Spurerkennung, eine Objekterkennung, eine Tiefenerkennung (3D-Schätzung der Bildbestandteile), semantische Erkennung, oder dergleichen mehr. Im Rahmen des Trainings werden die Ausgaben beider Ausgangsschnittstellen optimiert.

Ein Verfahren zum Korrigieren von Eingangs-Bilddaten einer Mehrzahl von Kameras eines Rundumsichtsystems, die durch Regen, Lichteinfall und/oder Schmutz beeinträchtigt sind, umfasst die Schritte:

a) von den Kameras erfasste Eingangs-Bilddaten, die durch Regen, Lichteinfall und/oder Schmutz beeinträchtigt sind, werden einem trainierten künstlichen neuronalen Netzwerk bereitgestellt,

b) das trainierte künstliche neuronale Netzwerk ist dazu konfiguriert, die Eingangs-Bilddaten mit Regen-, Lichteinfall und/oder Schmutzbeeinträchtigung in Ausgabe-Bilddaten ohne Beeinträchtigung umzuwandeln und ein Sicherheitsmaß c, welches abhängig vom Grad der Benetzung durch Wasser, Lichteinfall und/oder Verschmutzung für ein Bild bzw. jedes Bild der Eingangs-Bilddaten ist und charakterisiert (beim Einsatz des trainierten Netzwerks) die Sicherheit, dass eine Bildkorrektur durch das Netzwerk zutreffend ist, zu ermitteln, und

c) das trainierte künstliche neuronale Netzwerk ist dazu konfiguriert, die Ausgabe-Bilddaten und das ermittelte Sicherheitsmaß c auszugeben.

Die korrigierten Ausgabe-Bilddaten ermöglichen vorteilhafterweise eine bessere maschinelle Objekterkennung auf den Ausgabe-Bilddaten, z.B. eine herkömmliche Fahrspur-/Objekt- oder Verkehrszeichendetektion, oder ein verbessertes Stitching (Zusammensetzen der simultan erfassten Bilder der Kameras) und Anzeige der zusammengesetzten Bilddaten.

Gemäß einer Ausführungsform enthalten die Eingangs-Bilddaten mindestens eine Folge (Videosequenz) von aufeinander folgend erfassten Eingangsbildern der Kameras.

In einer Ausführungsform sind die Kameras fahrzeuggebundene Umfelderfassungskameras.

In einer Ausführungsform wird in Schritt a) zusätzlich ein Faktor d dem trainierten künstlichen neuronalen Netzwerk bereitgestellt und in Schritt b) die (Stärke bzw. der Grad der) Bildkorrektur bzw. Umwandlung in Abhängigkeit von dem Faktor d gesteuert.

Gemäß einer Ausführungsform wird der Faktor d geschätzt und bei der Schätzung wird die Beeinträchtigung der aktuellen erfassten Eingangs-Bilddaten berücksichtigt. Kumulativ oder alternativ kann die Schätzung des Faktors d der aktuell erfassten

Eingangs-Bilddaten den oder die Faktor(en) d der vorhergehend erfassten Bilddaten berücksichtigen.

- 5 Gemäß einer Ausführungsform kann bei der Bestimmung bzw. Schätzung des Faktors d eine zeitliche Entwicklung des Faktors d berücksichtigt werden. Hierzu gehen in die Schätzung die zeitliche Entwicklung des Faktors d und eine Sequenz von Eingangsbildern ein. Information über die zeitliche Entwicklung der Helligkeit kann auch für Bildregionen mit unterschiedlichen Faktoren d herangezogen werden.
- 10 In einer Ausführungsform wird für jede der Kameras des Rundumsichtsystems ein eigener Faktor d geschätzt bzw. bestimmt. Dies ermöglicht die individuelle Umwandlung für Bilddaten der einzelnen (Fahrzeug-) Kameras insbesondere in Abhängigkeit von der aktuellen Beeinträchtigung des Bildes der jeweiligen Kamera.
- 15 Gemäß einer Ausführungsform mit einer fahrzeuggebundenen Umfelderkassungskamera werden bei der Bestimmung des Faktors d Informationen über die aktuelle Umgebungssituation des Fahrzeugs berücksichtigt. Informationen über die aktuelle Umgebungssituation können beispielsweise Regensensordaten, externe (V2X-Daten bzw. Daten eines Navigationssystems, z.B. GPS-Empfänger mit
- 20 digitaler Karte) orts aufgelöste Wetter- und/oder Sonnenstands Informationen, Fahrsituationsinformationen (Landstraße, Stadt, Autobahn, Tunnel, Unterführung) umfassen. Diese Informationen können (zumindest teilweise) auch über eine Bildverarbeitung aus den Kamerabilddaten gewonnen werden.
- 25 Beispielsweise kann der aktuelle Faktor d basierend auf Umgebungssituationsinformationen und aus der zeitlichen Reihenfolge von Bildern als auch aus der Historie des Faktors d geschätzt werden.
- 30 Die Schätzung des Faktors d kann somit beim Einsatz eines trainierten künstlichen neuronalen Netzwerks dynamisch erfolgen.
- 35 In einer Ausführungsform erfolgt eine Ausgabe der korrigierten Bilddaten der fahrzeuggebundenen Umfelderkassungskamera und des ermittelten Sicherheitsmaßes (bzw. der ermittelten Sicherheitsmaße) c und optional auch des Faktors d an mindestens eine ADAS-Detektionsfunktion, welche ADAS-relevante Detektionen ermittelt und ausgibt. ADAS-Detektionsfunktionen können bekannte Kanten- bzw.

Mustererkennungsverfahren umfassen sowie Erkennungsverfahren, die mittels eines künstlichen neuronalen Netzwerks relevante Bildobjekte erkennen und optional klassifizieren können.

5 In einer alternativen Ausführungsform kann der Ansatz erweitert werden und das künstliche neuronale Netzwerk zur Korrektur der Bilddaten kann mit einem neuronalen Netzwerk für ADAS-Detektionsfunktionen, z.B. Spurerkennung, Objekterkennung, Tiefenerkennung, semantische Erkennung, kombiniert werden. Somit wird kaum zusätzlicher Aufwand in der Rechenzeit verursacht. Nach dem Training kann die (erste)
10 Ausgangsschnittstelle für die Ausgabe der umgewandelten (korrigierten) Bilddaten eliminiert werden, so dass bei Einsatz im Fahrzeug nur die (zweite) Ausgangsschnittstelle für die ADAS-Detektionen vorhanden ist.

15 In einer weiteren Ausführungsform kann das gelernte Verfahren statt einer Rekonstruktion von unklaren bzw. beeinträchtigten Bilddaten umgekehrt genutzt werden, um aus dem gelernten Rekonstruktionsprofil in aufgenommenen Bilddaten künstlich Regen oder Schmutz für eine Simulation zur Absicherung hinzuzufügen.

20 In einer weiteren Ausführungsform kann das gelernte Rekonstruktionsprofil auch zur Bewertung der Güte einer künstlichen Regensimulation in aufgenommenen Bilddaten herangezogen werden.

25 Gemäß einer weiteren Ausführungsform kann das Verfahren bei Augmented Reality und im Bereich Dash-Cam und Unfall Recordings angewendet werden.

30 Die Erfindung betrifft weiterhin eine Vorrichtung mit mindestens einer Datenverarbeitungseinheit konfiguriert zum Korrigieren von Eingangs-Bilddaten einer Mehrzahl von Kameras eines Rundumsichtsystems, die durch Regen, Lichteinfall und/oder Schmutz beeinträchtigt sind, in Ausgabe-Bilddaten. Die Vorrichtung umfasst: eine Eingangsschnittstelle, ein trainiertes künstliches neuronales Netzwerk und eine (erste) Ausgangsschnittstelle.

35 Die Eingangsschnittstelle ist konfiguriert zum Empfangen von Eingangs-Bilddaten, die durch Regen, Lichteinfall und/oder Schmutz beeinträchtigt sind, die von den Kameras erfasst worden sind. Das trainierte künstliche neuronale Netzwerk ist dazu konfiguriert, die beeinträchtigten Eingangs-Bilddaten in Ausgabe-Bilddaten ohne Beeinträchtigung

umzuwandeln und ein Sicherheitsmaß c auszugeben, welches abhängig vom Grad der Benetzung durch Wasser, Lichteinfall und/oder Verschmutzung für ein Bild bzw. jedes Bild der Eingangsbilddaten ist und ein Maß für die Sicherheit des Netzwerkes in seine berechnete Ausgabe ist bzw. die Sicherheit des Netzwerks charakterisiert, dass die Bildkorrektur des Netzwerks bzw. durch das Netzwerk zutreffend ist.

Die (erste) Ausgangsschnittstelle ist dazu konfiguriert, die umgewandelten (korrigierten) Bilddaten und das bzw. die ermittelte(n) Sicherheitsmaß(e) c auszugeben.

Gemäß einer Ausführungsform enthalten die Eingangsbilddaten mindestens eine Folge von aufeinander folgend erfassten Eingangsbildern als Eingangs-Bilddaten und das künstliche neuronale Netzwerk ist anhand von mindestens einer Folge von aufeinander folgenden Eingangs- und Ausgangsbildern als Bilddaten trainiert worden.

Die Vorrichtung bzw. die Datenverarbeitungseinheit kann insbesondere einen Mikrocontroller oder –prozessor, eine Zentrale Verarbeitungseinheit (CPU), ein Grafische Verarbeitungseinheit (GPU), einen Digital Signal Processor (DSP), einen ASIC (Application Specific Integrated Circuit), einen FPGA (Field Programmable Gate Array) und dergleichen mehr sowie Software zur Durchführung der entsprechenden Verfahrensschritte umfassen.

Gemäß einer Ausführungsform ist die Datenverarbeitungseinheit in eine hardwarebasierter Bildvorverarbeitungsstufe (Image Signal Processor, ISP) implementiert.

In einer Ausführungsform ist das trainierte künstliche neuronale Netzwerk zur Bildkorrektur ein Bestandteil eines fahrzeugseitigen ADAS-Detektions-Neuronalen-Netzwerks, z.B. zur semantischen Segmentierung, Fahrspurdetektion bzw. Objektdetektion, mit einer geteilten Eingangsschnittstelle (Eingangs- bzw. Merkmalsrepräsentationsschichten), und zwei separaten Ausgangsschnittstellen (Ausgangsschichten) ist, wobei die erste Ausgangsschnittstelle zur Ausgabe der umgewandelten Ausgabebilddaten und die zweite Ausgangsschnittstelle zur Ausgabe der ADAS-Detektionen (Bilderkennungsdaten) konfiguriert sind.

Die Erfindung betrifft weiterhin ein Computer Programmelement, welches, wenn damit eine Datenverarbeitungseinheit programmiert wird, die Datenverarbeitungseinheit dazu

anweist, ein Verfahren zur Bildkorrektur von Eingangs-Bilddaten einer Mehrzahl von Kameras eines Rundumsichtsystems in Ausgabe-Bilddaten durchzuführen.

5 Die Erfindung betrifft weiterhin ein Computerlesbares Speichermedium, auf dem ein solches Programmelement gespeichert ist.

10 Die Erfindung betrifft weiterhin die Verwendung eines Verfahrens zum maschinellen Lernen einer Bildkorrektur von Eingangs-Bilddaten einer Mehrzahl von Kameras eines Rundumsichtsystems in Ausgabe-Bilddaten zum Trainieren eines künstlichen neuronalen Netzwerks einer Vorrichtung mit mindestens einer Datenverarbeitungseinheit.

15 Die vorliegende Erfindung kann somit in digitalen elektronischen Schaltkreisen, Computer-Hardware, Firmware oder Software implementiert sein.

Wesentliche Vorteile sind:

- Ermöglichung einer Objekterkennung bei Beeinträchtigung (z.B. Beschlag) von einzelnen, mehreren oder allen Kameras des Rundumsichtsystems
- Verbesserung der Bildqualität bei seitlichen Satelliten- Rückfahrkameras zu 20 Anzeigezwecken, wenn das Bild bei Fahrt durch Wasser/ Schmutz gestört ist.
- Generierung eines Bilddatenstromes für Human und Computer Vision aus einem neuronalen Netzwerk für eine optimierte Korrespondenzsuche (Suche nach Feature-Übereinstimmung).

25 Neben dem Einsatz in Kraftfahrzeugen gibt es mannigfaltige Anwendungsgebiete:

- Assistenzsysteme in Bussen, Bahnen, Flugzeugen sowie Robotiksysteme
- Alle Anwendungen von Feature-basierter Detektion, z.B. auf optischem Fluss, Structure-from-Motion etc. basierende Detektionsverfahren, die bei Verschmutzung/Regen-/Lichtbeeinträchtigungen eine dramatische Degradation durch 30 fehlende Features erfahren
- somit insbesondere Assistenzsysteme, welche auf optischem Fluss zur Feature-Suche basieren
- Online-Kalibrierung von Kameras, welche bei Benetzung oder Verschmutzung aufgrund fehlender Features oder geringem Kontrast ebenfalls dramatisch an 35 Performance verliert.

Im Folgenden werden Ausführungsbeispiele und Fig. näher beschrieben. Dabei zeigen

Fig. 1: eine erste schematische Darstellung einer erfindungsgemäßen Vorrichtung
5 in einer Ausgestaltung;

Fig. 2: eine zweite schematische Darstellung einer erfindungsgemäßen Vorrichtung
in einer Ausgestaltung in einem Fahrzeug;

Fig. 3: ein System zur Korrektur von beeinträchtigten Kamerabildern;

Fig. 4: ein System mit einem ersten neuronalen Netz zur Bildkorrektur;

10 Fig. 5: ein System mit kombinierter Bildkorrektur und Detektionsfunktionen; und

Fig. 6 ein abgewandeltes System, bei dem die Bildkorrektur nur im Rahmen des
Trainings berechnet und ausgegeben wird.

Wie man aus Fig. 1 erkennen kann, kann eine erfindungsgemäße Vorrichtung 1 zur
15 Bildkorrektur von Eingangs-Bilddaten von mehreren Kameras eines
Rundumsichtsystems mehrere Einheiten bzw. Schaltungskomponenten aufweisen. Bei
dem in Fig. 1 dargestellten Ausführungsbeispiel weist die Vorrichtung 1 zur
Bildkorrektur mehrere Fahrzeugkameras 2-i auf, die jeweils Kamerabilder bzw.
Videodaten erzeugen. Bei dem in Fig. 1 dargestellten Ausführungsbeispiel weist die
20 Vorrichtung 1 vier Fahrzeugkameras 2-i zur Erzeugung von Kamerabildern auf. Die
Anzahl der Fahrzeugkameras 2-i kann für unterschiedliche Anwendungen variieren. Die
erfindungsgemäße Vorrichtung 1 weist mindestens zwei Fahrzeugkameras zur
Erzeugung von Kamerabildern auf. Die Kamerabilder von benachbarten
Fahrzeugkameras 2-i weisen typischerweise überlappende Bildbereiche auf.

25 Die Vorrichtung 1 enthält eine Datenverarbeitungseinheit 3, welche die von den
Fahrzeugkameras 2-i erzeugten Kamerabilder zu einem zusammengesetzten
Gesamtbild zusammensetzt. Wie in Fig. 1 dargestellt, weist die
Datenverarbeitungseinheit 3 ein System zur Bildkorrektur bzw. zur Bildumwandlung 4
30 auf. Das System zur Bildumwandlung 4 erzeugt aus den Eingangsbilddaten (Ini) der
Fahrzeugkameras (2-i), die zumindest teilweise durch Regen, Lichteinfall und/oder
Schmutz beeinträchtigt sind, korrigierte Ausgangs- bzw. Ausgabebilddaten (Opti) ohne
Beeinträchtigung. Die optimierten Ausgabebilddaten der einzelnen Fahrzeugkameras
2-i werden zu einem zusammengesetzten Gesamtbild zusammengesetzt (sogenanntes
35 Stitching). Das von der Bildverarbeitungseinheit 3 aus den korrigierten Bilddaten (Opti)
zusammengesetzte Gesamtbild wird anschließend durch eine Anzeigeeinheit 5 einem

Nutzer angezeigt. Bei einer möglichen Ausführungsform wird das System zur Bildkorrektur 4 durch eine eigenständige Hardwareschaltung gebildet, welche die Bildkorrektur vornimmt. Bei einer alternativen Ausführungsform führt das System Programmbefehle bei Durchführung eines Verfahrens zur Bildkorrektur aus.

5

Die Datenverarbeitungseinheit 3 kann einen oder mehrere Bildverarbeitungsprozessoren aufweisen, wobei sie die von den verschiedenen Fahrzeugkameras 2-i empfangenen Kamerabilder bzw. Videodaten umwandelt und anschließend zu einem zusammengesetzten Gesamtbild zusammensetzt (Stitching).

10

Bei einer möglichen Ausführungsform wird das System zur Bildumwandlung 4 durch einen dafür vorgesehenen Prozessor gebildet, welcher die Bildkorrektur parallel zu dem oder den übrigen Prozessoren der Datenverarbeitungseinheit 3 vornimmt. Durch die parallele Datenverarbeitung wird die benötigte Zeit zur Verarbeitung der Bilddaten gesenkt.

15

Figur 2 zeigt eine weitere schematische Darstellung einer erfindungsgemäßen Vorrichtung 1 in einer Ausgestaltung. Die in Fig. 2 dargestellte Vorrichtung 1 wird in einem Surround View System eines Fahrzeugs 10, insbesondere eines Personenkraftwagens oder eines Lastkraftwagens, eingesetzt. Hierbei können sich die vier verschiedenen Fahrzeugkameras 2-1, 2-2, 2-3, 2-4 an verschiedenen Seiten des Fahrzeugs 10 befinden und weisen entsprechende Sichtbereiche (gestrichelte Linien) vor V, hinter H, links L und rechts R dem bzw. des Fahrzeug(s) 10 auf.

20

Beispielsweise befindet sich die erste Fahrzeugkamera 2-1 an einer Vorderseite des Fahrzeugs 10, die zweite Fahrzeugkamera 2-2 an einer Hinterseite des Fahrzeugs 10, die dritte Fahrzeugkamera 2-3 an der linken Seite des Fahrzeugs 10 und die vierte Fahrzeugkamera 2-4 an der rechten Seite des Fahrzeugs 10. Die Kamerabilder von zwei benachbarten Fahrzeugkameras 2-i weisen überlappende Bildbereiche VL, VR, HL, HR auf. Bei den Fahrzeugkameras 2-i handelt es sich bei einer möglichen Ausführungsform um sogenannte Fischaugenkameras, die einen Blickwinkel von mindestens 185° aufweisen. Die Fahrzeugkameras 2-i können die Kamerabilder bzw. Kamera-Bild-Frames bzw. Videodaten bei einer möglichen Ausführungsform über eine Ethernet-Verbindung an die Datenverarbeitungseinheit 3 übertragen. Die Datenverarbeitungseinheit 3 berechnet aus den Kamerabildern der Fahrzeugkameras 2-i ein zusammengesetztes Surround-View-Kamerabild, das auf der Anzeige 5 des Fahrzeugs 10 dem Fahrer und/oder einem Passagier angezeigt wird. In manchen Fällen weichen die Sichtverhältnisse einer Kamera, z.B. der hinteren Fahrzeugkamera

25

30

35

2-2 von denen der übrigen Kameras 2-1, 2-3, 2-4 ab, weil die Linse der hinteren Fahrzeugkamera 2-2 mit Regentropfen benetzt oder verschmutzt ist.

Bei einem gleichzeitigen bzw. gemeinsamen Training eines künstlichen neuronalen Netzwerks mit beeinträchtigten (beispielsweise der Rearviewkamera 2-2) und unbeeinträchtigten Bildern (beispielsweise der Front- 2-1 und Seitenkameras 2-3, 2-4) lernt das neuronale Netz optimale Parameter für die Bildkorrektur in dieser Situation.

Bei dem gemeinsamen Training für mehrere Fahrzeugkameras 2-i werden in einer ersten Anwendung bevorzugt Groundtruth-Daten verwendet, welche eine für alle Zielkameras 2-1, 2-2, 2-3, 2-4 angewandte Bildqualität ohne Beeinträchtigung durch Regen, Lichteinfall oder Schmutz aufweisen. Mit diesen Groundtruth-Daten als Referenz und den Eingangsdaten der Zielkameras 2-1, 2-2, 2-3, 2-4, die derartige Beeinträchtigungen aufweisen können, wird ein neuronales Netz CNN1, CNN10, CNN11, CNN12, hinsichtlich eines optimalen Parametersets für das Netz trainiert.

In einer weiteren Anwendung kann das neuronale Netz für die gemeinsamen Kameras 2-i dahingehend trainiert werden, dass auch im Falle von fehlenden Trainingsdaten und Groundtruth-Daten für eine Kamera, beispielsweise einer Seitenkamera 2-3 oder 2-4, das Netz die Parameter für diese Kamera 2-3 oder 2-4 mit den fehlenden Daten basierend auf den Trainingsdaten der anderen Kameras 2-1, 2-2 und 2-4 oder 2-3 trainiert und optimiert.

In einem letzten Beispiel nutzt das neuronale Netz zeitlich unterschiedliche und mit den einzelnen Kameras 2-i korrelierte Trainings- und Groundtruth-Daten, welche durch die verschiedenen Kameras 2-i zu unterschiedlichen Zeitpunkten erfasst bzw. aufgenommen wurde. Dazu können Informationen von Features oder Objekten und deren Groundtruth-Daten verwendet werden, welche beispielsweise zu einem Zeitpunkt t von der Frontkamera 2-1 und zu einem Zeitpunkt $t+n$ von den Seitenkameras 2-3, 2-4 aufgenommen wurden. Diese Feature oder Objekte und deren Groundtruth-Daten können fehlende Informationen in den Training- und Groundtruth-Daten der jeweils anderen Kameras ersetzen, wenn sie in den Bildern der anderen Kameras 2-i und dann vom Netz als Trainingsdaten verwendet werden. Auf diese Weise kann das Netz die Parameter für alle Seitenkameras 2-3, 2-4 optimieren, und gegebenenfalls fehlende Informationen in den Trainingsdaten kompensieren.

Dies führt bei der Verwendung von mehreren Fahrzeugkameras 2-i zu einer angepassten Bildkorrektur für alle Fahrzeugkameras 2-i, da die individuellen Beeinträchtigungen im Gesamtverbund explizit erfasst und trainiert werden.

5 Fig. 3 zeigt schematisch eine allgemeine Übersicht eines Systems zur Korrektur von Kamerabildern bei Regen, Lichteinfall und/oder Verschmutzung. Wesentlicher Bestandteil ist ein künstliches neuronales Netzwerk CNN1, welches in einer Trainingsphase lernt, einem Satz von Trainings-Eingangsbildern In (In1, In2, In3, ...) einen Satz von korrespondierenden korrigierten Trainings-(Soll-)Ausgabebildern Out (Out1, Out2, Out3, ...) zuzuordnen. Zuordnen heißt hierbei, dass das neuronale Netz CNN1
10 lernt, ein korrigiertes Bild zu generieren. Ein Eingangsbild (In1, In2, In3, ...) kann z.B. eine Straßenszene bei Regen enthalten, auf dem mit dem menschlichen Auge nur unscharfe bzw. verschwommene größere Objekte wie z.B. eine große Fahrbahnmarkierung, die ein Fahrrad darstellt, und der Himmel erkennbar sind. Auf dem korrespondierenden korrigierten Bild (Out1, Out2, Out3, ...) sind zusätzlich die
15 Konturen einer Straßenkreuzung, eine Ampel, ein Baum und eine Straßenlaterne erkennbar. Realistische photographische Abbildungen für Farbbilder mit und ohne Beeinträchtigungen durch Regentropfen sind beispielsweise in der Abbildung Fig. 1 von Porav et al. zu sehen. Der Begriff „ein Eingangsbild“ kann hierbei zeitgleich erfasste
20 Eingangsbilder von mehreren bzw. allen Einzelkameras 2-i bedeuten, da ja eine Mehrzahl von Kameras 2-i vorliegt. Entsprechend kann der Begriff „ein Ausgangsbild“ die Soll-Ausgabebilder für mehrere bzw. alle Einzelkameras 2-i beinhalten.

Optional dient ein Faktor d als zusätzliche Eingangsgröße für das neuronale Netz
25 CNN1. Der Faktor d ist ein Steuerungsparameter, der den Grad der Korrektur der Beeinträchtigung (Regen, Lichteinfall bzw. Schmutz) des Bildes steuert. Beim Training kann der Faktor d für ein Bildpaar aus Trainingsbild und korrigiertem Bild (In1, Out1; In2, Out2; In3, Out3; ...) vorab oder im Rahmen des Trainings aus dem Bildpaar (In1, Out1; In2, Out2; In3, Out3; ...) ermittelt und dem neuronalen Netz CNN1 bereitgestellt
30 werden. Dadurch kann der Faktor d mitgelernt werden.

Beim Einsatz des trainierten neuronalen Netzes CNN1 kann über eine Vorgabe eines Faktors d gesteuert werden, wie stark das neuronale Netz CNN1 ein aktuell erfasstes
35 Bild korrigiert – man kann sich den Faktor d auch als einen externen Regressionsparameter vorstellen (mit einer beliebigen Abstufung). Da der Faktor d möglichen Schwankungen im Bereich von +/- 10% unterliegen kann, wird dies während

des Trainings berücksichtigt. Der Faktor d kann während des Trainings um circa $\pm 10\%$ verrauscht werden (z.B., während der verschiedenen Epochen des Trainings des neuronalen Netzwerkes), um während der Inferenz im Fahrzeug robust gegenüber Fehlschätzungen des Faktors d im Bereich von ca. $\pm 10\%$ zu sein. Mit anderen Worten
5 liegt die nötige Genauigkeit vom Faktor d im Bereich von $\pm 10\%$ - somit ist das neuronale Netz CNN1 robust gegenüber Abweichungen bei Schätzungen von diesem Parameter.

Alternativ oder ergänzend kann der Faktor d für eine erfolgte Bildkorrektur durch das
10 trainierte neuronale Netzwerk CNN1 ausgegeben werden. Dadurch erhalten nachgelagerte Bilderkennungs- oder Bildanzeigefunktionen eine Information darüber, wie stark das ursprünglich erfasste Bild korrigiert wurde.

Das künstliche neuronale Netzwerk CNN1 ist derart ausgelegt, dass es ein
15 Sicherheitsmaß c ermittelt, das vom Grad der Benetzung durch Wasser, Lichteinfall und/oder Verschmutzung für ein Eingangs-Bild abhängt. Die Auslegung des Netzwerkes kann beispielsweise durch ein entsprechendes Design der Architektur des künstlichen neuronalen Netzwerkes CNN1 erfolgen. Nach Abschluss des maschinellen Lernens kann das künstliche neuronale Netzwerk CNN1 das Sicherheitsmaß c für ein neues
20 Eingangsbild bestimmen und ausgeben. Das Sicherheitsmaß c hängt also vom Grad der Beeinträchtigung durch eine Benetzung mit Regen bzw. Wasser, durch Lichteinfall und/oder durch Verschmutzung ab und charakterisiert beim Einsatz des trainierten Netzwerkes die Sicherheit, dass eine Bildkorrektur zutreffend ist. Sofern der Begriff „ein Eingangsbild“ die zeitgleich erfassten Eingangsbilder von mehreren bzw. allen
25 Einzelkameras 2-i bedeutet, kann der Begriff „ein Sicherheitsmaß c “ auch bedeuten, dass für jedes der verschiedenen zeitgleich erfassten Eingangsbilder ein eigenes (möglicherweise unterschiedliches) Sicherheitsmaß c ermittelt wird.

In Fig. 3 sind drei Bildpaare $In_1 + Out_1$, $In_2 + Out_2$, $In_3 + Out_3$ schematisch dargestellt.
30 Entsprechend wird durch das neuronale Netzwerk CNN1 dahingehend trainiert bzw. ist derart designt worden, dass es für jedes Eingangsbild eines Bildpaars ein Sicherheitsmaß c_1 , c_2 bzw. c_3 ermitteln und ausgeben kann.

Da die Qualität der Korrektur von Bildern von Regen, Lichteinfall oder Verschmutzung
35 abhängig ist von zahlreichen Faktoren (wie z.B. das Vorhandensein eines ähnlichen Falls in den Trainingsdaten, ausreichende Belichtung für eine mögliche Korrektur,

Vermeidung von mehrdeutigen Szenarien, etc.), gibt das Netzwerk neben der Korrektur des Bildes ein Maß für die Sicherheit aus, mit welcher das Netzwerk seine Entscheidung trifft. Dieses Sicherheitsmaß c kann eine der folgenden Umsetzungsformen oder eine Kombination dieser umfassen:

- Ein Konfidenzmaß c_{Prob} : Hierbei wird die Ausgabe des Netzwerkes derart kalibriert, dass diese probabilistisch interpretiert werden kann als Wahrscheinlichkeit, mit welcher das Netzwerk die richtige Entscheidung trifft. Werte hierfür sind auf einen Bereich zwischen $[0,1]$ normiert und diese entsprechen dem Spektrum von 0% Wahrscheinlichkeit bis 100% Wahrscheinlichkeit, dass das Netzwerk eine korrekte Korrektur eines Bildes berechnet hat. Diese Kalibrierung kann nach Abschluss des eigentlichen maschinellen Lernverfahrens anhand eines Trainingsbilddatensatzes durch die anschließende Überprüfung der Güte des Lernens anhand eines Validierungsbilddatensatzes erfolgen. Auch der Validierungsbilddatensatz enthält Bildpaare von jeweils einem ersten Bild mit Regen-, Lichteinfall und/oder Schmutzbeeinträchtigung und als korrespondierendes Soll-Ausgabe-Bild ein zweites Bild derselben Szene ohne Beeinträchtigung. In der Praxis kann ein Teil der Eingangs- und Soll-Ausgabebilder zurückbehalten, d.h. nicht für das maschinelle Lernverfahren verwendet, und anschließend zur Validierung verwendet werden.
- Ein Streuungsmaß ähnlich einer Standardabweichung c_{Dev} : Hierbei wird eine Unsicherheit der Netzwerkausgabe derart geschätzt, dass diese die Streuung des Netzwerkausgangs beschreibt. Dies kann unterschiedlich umgesetzt werden. Möglichkeiten sind hierfür die Unterteilung in Mess- und Modellunsicherheit. Die Messunsicherheit bezieht sich auf Unsicherheiten, die durch die Eingangsdaten verursacht werden, z.B. leichte Störungen. Diese können über einen weiteren Ausgang dem Netzwerk hinzugefügt werden und durch eine Änderung der Fehlerfunktion trainiert werden. Die Modellunsicherheit bezieht sich auf Unsicherheiten, die durch die begrenzte Abbildungsgenauigkeit und Generalisierbarkeit eines Netzwerkes verursacht werden. Dies bezieht sich auf Faktoren, wie Umfang der Trainingsdaten und Architektur des Netzwerkdesigns. Die Modellunsicherheit lässt sich schätzen, z.B., durch Monte Carlo Dropout oder Netzwerk-Ensembles. Die Modellunsicherheit und die Messunsicherheit können additiv zusammengefasst werden.
- Eine Kombination aus Konfidenz- und Streuungsmaß

Das Sicherheitsmaß c kann für das gesamte Bild, Bildbereiche oder die einzelnen Pixel des Bildes berechnet werden.

Basierend auf dem Sicherheitsmaß c können folgende Entscheidungen getroffen werden:

- 5 - c_Prob niedrig: Das Netzwerk hat eine niedrige Konfidenz in seine Schätzung – Fehlschätzungen treten vermehrt auf.
- c_Prob hoch: Das Netzwerk hat eine hohe Konfidenz in seine Schätzung – die Bildkorrektur ist in den meisten Fällen korrekt.
- 10 - c_Dev niedrig: Die Streuung der Bildkorrektur des Netzwerkes ist gering – somit prädiziert das Netzwerk eine sehr genaue Bildkorrektur.
- c_Dev hoch: Die geschätzte Streuung der Bildkorrektur ähnlich einer Standardabweichung ist hoch und die Ausgabe des Netzwerkes ist ungenauer/weniger scharf – geringe Änderung der Eingangsdaten oder in der Modellierung des Netzwerkes würden Abweichungen in der Bildkorrektur
- 15 verursachen.
- Kombinationen:
 - c_Prob hoch und c_Dev niedrig: eine sehr zuverlässige und genaue Bildkorrektur, die mit hoher Sicherheit angenommen werden kann
 - c_Prob niedrig und c_Dev hoch: eine sehr unsichere und ungenaue
 - 20 Bildkorrektur, die eher abgelehnt werden würde
 - c_Prob hoch und c_Dev hoch bzw. c_Prob niedrig und c_Dev niedrig: diese Korrekturen sind mit Unsicherheiten behaftet und hier ist eine vorsichtige Verwendung der Bildkorrekturen zu empfehlen

25 Die Hinzunahme der Sicherheitsmaße ist insbesondere für sicherheitskritische Funktionen relevant.

30 Eine Möglichkeit zur Erzeugung der Trainingsdaten (Trainingsbilder ($In_1, In_2, In_3, \dots$) und zugeordnete korrigierte Bilder ($Out_1, Out_2, Out_3, \dots$)) besteht in der Aufnahme von Bilddaten mit einem „Stereokameraaufbau“ wie in Porav et al. anhand der dortigen Fig. 8 beschrieben: vor zwei identischen Kameramodulen mit geringem Abstand zueinander wird eine zweigeteilte Kammer mit transparenten Scheiben angeordnet, die Kammer z.B. vor dem rechten Stereokameramodul wird mit Wassertropfen besprüht, während die Kammer vor dem linken Stereokameramodul von Beeinträchtigungen freigehalten wird.

Um eine Beeinträchtigung durch Lichteinfall in analoger Weise zu simulieren, kann z.B. eine Lichtquelle nur auf eine Kammer gerichtet werden. Oder im Fall von Schmutz kann dieser ebenfalls nur auf einer Kammer angebracht werden.

- 5 Alternativ kann man zur Erzeugung der Trainings-Bildpaare unbeeinträchtigte Bilder nehmen, und diese mit Renderingverfahren, welche Auswirkungen von Regen, Lichteinfall oder Schmutz im Bild simulieren, schlechtrechnen.

10 Wenn das neuronale Netz CNN1 trainiert ist, erfolgt eine Bildkorrektur nach folgendem Schema:

Eingangsbild $\rightarrow$ CNN1

Optional: Faktor $d \rightarrow$ CNN1

CNN1 $\rightarrow$ korrigiertes Ausgangs-/Ausgabebild + Sicherheitsmaß c .

- 15 Fig. 4 zeigt ein System mit einem trainierten neuronalen Netz CNN1 zur Bildkorrektur. Das trainierte neuronale Netz CNN1 erhält ursprüngliche Eingangsbilddaten (I_{ni}) der mehreren Kameras 2-i als Input. Optional kann ein Faktor d vorgegeben oder durch das neuronale Netz CNN1 anhand der Eingangsbilddaten (I_{ni}) bestimmt werden, der vorgibt (steuert), wie stark die Eingangsbilddaten korrigiert werden sollen. Das neuronale Netz
20 berechnet korrigierte Bilddaten (O_{pti}) der mehreren Kameras 2-i ohne Beeinträchtigungen und ein bzw. mehrere Sicherheitsmaß(e) c . Die korrigierten Bilddaten (O_{pti}) der mehreren Kameras 2-i und das mindestens eine Sicherheitsmaß c werden ausgegeben.

- 25 Die Fig. 5 und 6 zeigen Ausführungsbeispiele für mögliche Kombinationen eines ersten Netzwerks für die Bildkorrektur mit einem oder mehreren Netzwerken für (Detektions-) Funktionen für Fahrerassistenzsysteme und/oder Automatisiertes Fahren.

- 30 Fig. 5 zeigt ein neuronales Netzwerk CNN10 für die Bildkorrektur eines Eingangsbildes (I_{ni}) ggfs. gesteuert durch einen Faktor d , welches Merkmalsrepräsentationsschichten (als Eingangs- bzw. untere Schichten) mit einem Netzwerk für Detektionsfunktionen ($fn1$, $fn2$, $fn3$, $fn4$) teilt. Die Detektionsfunktionen ($fn1$, $fn2$, $fn3$, $fn4$) sind hierbei Bildverarbeitungsfunktionen, die für ADAS oder AD-Funktionen relevante Objekte, Strukturen, Eigenschaften (allgemein: Merkmale) in den Bilddaten detektieren. Viele
35 solcher Detektionsfunktionen ($fn1$, $fn2$, $fn3$, $fn4$), die auf maschinellem Lernen beruhen, sind bereits entwickelt oder Gegenstand aktueller Entwicklung (z.B:

Verkehrszeichenklassifikation, Objektklassifikation, semantische Segmentierung, Tiefenschätzung, Fahrspurmarkierungserkennung und -lokalisierung). Auf korrigierten Bildern (Opti) liefern Detektionsfunktionen (fn1, fn2, fn3, fn4) des zweiten neuronalen Netzes CNN2 bessere Ergebnisse als auf den ursprünglichen beeinträchtigten Eingangsbilddaten (Ini). In den Merkmalsrepräsentationsschichten des neuronalen Netzwerks CNN10 werden gemeinsame Merkmale für die Bildkorrektur und für die Detektionsfunktionen gelernt.

Das neuronale Netz CNN10 mit geteilten Eingangsschichten und zwei separaten Ausgängen weist einen ersten Ausgang CNN 11 zur Ausgabe der korrigierten Ausgangs-/Ausgabebilddaten (Opti) sowie einen zweiten Ausgang CNN 12 zur Ausgabe der Detektionen: Objekte, Tiefe, Spur, Semantik, etc. auf.

Dadurch, dass beim Training die Merkmalsrepräsentationsschichten sowohl hinsichtlich der Bildkorrektur als auch der Detektionsfunktionen (fn1, fn2, fn3, fn4) optimiert werden, bewirkt eine Optimierung der Bildkorrektur zugleich eine Verbesserung der Detektionsfunktionen (fn1, fn2, fn3, fn4).

Falls nun eine Ausgabe des korrigierten Bildes (Opti) nicht gewünscht bzw. nicht erforderlich ist, kann der Ansatz weiter variiert werden, wie anhand von Fig. 6 erläutert wird.

Fig. 6 zeigt einen auf dem System der Fig. 5 basierenden Ansatz zur neuronalen-Netz-basierten Bildkorrektur durch Optimierung der Merkmale. Um Rechenzeit zu sparen, werden die Merkmale für die Detektionsfunktionen (fn1, fn2, fn3, fn4) während des Trainings hinsichtlich Bildkorrektur und hinsichtlich der Detektionsfunktionen (fn1, fn2, fn3, fn4) optimiert.

Zur Laufzeit, d.h. beim Einsatz des trainierten neuronalen Netzes (CNN10, CNN11, CNN12), werden keine korrigierten Bilder (Opti) berechnet.

Dennoch sind die Detektionsfunktionen (fn1, fn2, fn3, fn4) - wie bereits erläutert - durch das gemeinsame Training von Bildkorrektur und Detektionsfunktionen verbessert verglichen mit einem System mit nur einem neuronalen Netz (CNN2) für Detektionsfunktionen (fn1, fn2, fn3, fn4), bei dem im Training auch nur die Detektionsfunktionen (fn1, fn2, fn3, fn4) optimiert worden sind.

In der Trainingsphase wird durch eine zusätzliche Ausgangsschnittstelle (CNN11) das korrigierte Bild (Opti) ausgegeben und mit der Ground Truth (dem korrespondierenden

korrigierten Trainingsbild) verglichen. In der Testphase bzw. zur Laufzeit kann dieser Ausgang (CNN11) weiterverwendet werden oder, um Rechenzeit zu sparen, abgeschnitten werden. Die Gewichte für die Detektionsfunktionen (fn1, fn2, fn3, fn4) werden bei diesem Training mit dem zusätzlichen Ausgang (CNN11) dahingehend modifiziert, dass sie die Bildkorrekturen für die Detektionsfunktionen (fn1, fn2, fn3, fn4) berücksichtigen. Die Gewichte der Detektionsfunktionen (fn1, fn2, fn3, fn4) lernen somit implizit die Information über die Helligkeitsverbesserung.

Weitere Aspekte und Ausführungsformen eines Assistenzsystems, welches die Bilddaten des zugrundeliegenden Kamerasystems trotz Beeinträchtigung durch Regen, Lichteinfall oder Schmutz algorithmisch in eine Darstellung umrechnet, welche einer Aufnahme ohne diese Beeinträchtigungen entspricht, werden im Folgenden dargelegt. Das umgerechnete Bild kann dann entweder zu reinen Anzeigezwecken oder als Input für Feature-basierte Erkennungsalgorithmen dienen.

1) In einer ersten Ausführung basiert die Berechnung in einem System beispielsweise auf einem neuronalen Netz, welches einer Detektions- oder Anzeigeeinheit vorgelagert ein Eingangsbild mit Beschlag, Schmutz oder Wassertropfen mit wenig Kontrast und Farbinformationen in eine bereinigte Repräsentation umwandelt. Für diese Aufgabe wurde das neuronale Netz mit einem Datensatz bestehend aus „beschlagenen Eingangsbildern“ und den dazugehörigen „bereinigten Bildern“ trainiert.

2) Insbesondere wird durch die Verwendung von bereinigten Bildern das neuronale Netz dahingehend trainiert, dass in den zu verbessernden Bildpaaren vorkommende Features trotz Beschlag oder Schmutz für eine spätere Korrespondenzsuche oder Objekterkennung erhalten und bestenfalls sogar verstärkt werden.

3) Beim Trainieren des Netzes zur Bildkorrektur/-verbesserung können Feature-basierte Verfahren zur Anzeige und Objektdetektion berücksichtigt werden, so dass das Verfahren für die zu erkennenden Features spezialisiert wird und diese Features explizit für eine nachfolgende Bearbeitung hervorhebt.

4) In einer weiteren Ausführung kann das Verfahren zur Bildverbesserung bzw. -korrektur in einer hardwarebasierten Bildvorverarbeitungsstufe, der ISP, integriert sein. Diese ISP wird hardwareseitig um ein neuronales Netz ergänzt, welches die

Umwandlung vornimmt und die aufbereiteten Informationen mit den Originaldaten möglichen Detektions- oder Anzeigeverfahren zur Verfügung stellt.

5) In einer weiteren Ausführung kann das System mit dem neuronalen Netz dahingehend trainiert werden, dass es zusätzliche Informationen von Beschlag freien Kameras wie Seitenkameras verwendet, um die Umrechnung für die beschlagene Bereiche weiter zu verbessern. Hier wird das Netz dann weniger mit Einzelbildern für jede Kamera individuell trainiert, sondern als Gesamtsystem bestehend aus mehreren Kamerasystemen.

6) In einer weiteren Anwendung können dem Netz zum Trainieren neben Verschmutzungs- oder Beschlaginformationen Informationen zur Bildqualität zur Verfügung gestellt werden. Dahingehend kann das System und das Verfahren dahingehend optimiert werden, dass es für eine Objekterkennung und Human Vision optimierte Bilddaten berechnet.

7) In weiteren Anwendungen sei der Verschmutzungsgrad der Kameras unterschiedlich. So sei beispielhaft eine seitlich am Fahrzeug angebrachte Satellitenkamera stärker verschmutzt als eine vorne am Fahrzeug angebrachte Satellitenkamera. So sei in diesem Fall das künstliche neuronale Netz so ausgelegt, trainiert und optimiert, dass es beispielsweise die Bildinformationen und Bildeigenschaften von Satellitenkameras ohne Beschlag zur Berechnung einer beschlagsfreien Darstellung in Bildern von Kameras mit Beschlag nutzt. Das daraus berechnete Bild kann dann zu Anzeigezwecken, aber auch zur Erkennung von Features genutzt werden. In den nachfolgenden Beispielen dienen die korrigierten Bildern von beschlagenen Kameras sowohl der Erkennung von Features für Optischen Fluss oder Structure from Motion sowie für Anzeigezwecke.

8) In einer Anwendung sei das Verfahren derart ausgelegt, dass es mit einem gemeinsamen Training eines künstlichen neuronalen Netzes mit Bildern unterschiedlichen Verschmutzungsgrades (beispielsweise Beschlag auf den Seitenkameras) und klaren Bildern (beispielsweise für die Front- Rearview-kameras) optimale Parameter für alle Satellitenkameras gleichzeitig lernt.

Bei dem gemeinsamen Training für mehrere Fahrzeugkameras werden in einer ersten Anwendung bevorzugt Groundtruth Daten verwendet, welche eine für alle Zielkameras

angewandte Bildqualität aufweisen. Mit anderen Worten sind die Groundtruth Daten für alle Zielkameras derart ausbalanciert, dass beispielsweise bei einer Surroundviewanwendung keine Helligkeitsunterschiede in den Groundtruthdaten erkennbar ist. Mit diesen Groundtruthdaten als Referenz und den Eingangsdaten der Zielkameras, die unterschiedliche Beschlag haben können, wird ein neuronale Netz hinsichtlich eines optimalen Parametersets für ein Netz trainiert. Denkbar sind ebenfalls Daten mit unterschiedlich ausgeleuchteten Seitenbereichen, beispielsweise wenn sich das Fahrzeug neben einer Straßenlaterne befindet, oder das Fahrzeug auf einer Seite eine zusätzliche Lichtquelle aufweist.

9) In einer weiteren Anwendung kann das Netz für die gemeinsamen Kameras dahingehend trainiert werden, dass auch im Falle von fehlenden Trainingsdaten und Groundtruth Daten für eine Kamera, beispielsweise einer Seitenkamera, das Netz die Parameter für die Kamera mit den fehlenden Daten basierend auf den Trainingsdaten der anderen Kameras trainiert und optimiert. Dies kann beispielsweise als Einschränkung im Training des Netzes erreicht werden, beispielsweise als Annahme, dass die Korrektur und das Training aufgrund ähnlicher Bildqualität der Seitenkameras immer gleich sein muss.

10) In einem letzten Beispiel nutzt das Netz zeitlich unterschiedliche und mit den Kameras korrelierte Trainings- und Groundtruth Daten, welche durch die verschiedenen Kameras zu unterschiedlichen Zeitpunkten aufgenommen wurden. Dazu können Informationen von Features und deren Groundtruth Daten verwendet werden, welche beispielsweise zu einem Zeitpunkt t von der Frontkamera und zu einem Zeitpunkt $t+n$ von den Seitenkameras aufgenommen wurden. Diese Feature und deren Groundtruth Daten können fehlende Informationen in den Training- und Groundtruthdaten der jeweils anderen Kameras ersetzen, wenn sie in den Bildern der anderen Kameras und dann vom Netz als Trainingsdaten verwendet werden. Auf diese Weise kann das Netz die Parameter für alle Seitenkameras optimieren, und gegebenenfalls fehlende Informationen in den Trainingsdaten kompensieren.

Dies führt bei der Verwendung von mehreren Fahrzeugkameras 2-i zu einer angepassten Bildqualität für alle Fahrzeugkameras 2-i, da die individuellen Beschlagsprofile im Außenraum im Gesamtverbund explizit erfasst und trainiert werden.

- 11) Neben der Korrektur von Bildern bei Beschlag oder Wassertropfen erkennt das System Wassertropfen oder Schmutz für beispielsweise eine Aktivierung eines Scheibenwischers oder der Anzeige zur Aufforderung der Reinigung einer Satellitenkamera. Zusammen mit einer Helligkeitsdetektion lässt sich auf diese Weise
- 5 neben der Korrektur der Bilder eine Funktion der Rain Light Detection realisieren.

Patentansprüche

1. Verfahren zum maschinellen Lernen einer Bildkorrektur von Eingangs-Bilddaten einer Mehrzahl von Kameras (2-i) eines Rundumsichtsystems, wobei die
5 Eingangs-Bilddaten zumindest teilweise durch Regen, Lichteinfall und/oder Schmutz beeinträchtigt sind, in korrigierte Ausgabe-Bilddaten mittels eines künstlichen neuronalen Netzwerks (CNN1, CNN10, CNN11, CNN12), wobei das Lernen mit einer Vielzahl von Trainings-Bildpaaren (In1, Out1; In2, Out2; In3, Out3; ...) derart erfolgt, dass am Eingang des künstlichen neuronalen Netzwerks (CNN1, CNN10) jeweils ein
10 erstes Bild (In1, In2, In3, ...) mit Regen-, Lichteinfall und/oder Schmutzbeeinträchtigung und als Soll-Ausgabe-Bild ein zweites Bild (Out1, Out2, Out3, ...) derselben Szene ohne Beeinträchtigung bereitgestellt wird,
wobei das künstliche neuronale Netzwerk (CNN1, CNN10, CNN11, CNN12) derart ausgelegt ist, dass es ein Sicherheitsmaß c ermittelt, das vom Grad der Benetzung
15 durch Wasser, Lichteinfall und/oder Verschmutzung für ein Eingangs-Bild abhängt und das künstliche neuronale Netzwerk (CNN1, CNN10, CNN11, CNN12) nach Abschluss des maschinellen Lernens das Sicherheitsmaß c für ein neues Eingangsbild bestimmen und ausgeben kann.
2. Verfahren nach Anspruch 1, wobei die Trainings-Bildpaare (In1, Out1; In2, Out2; In3, Out3; ...) erzeugt werden, indem jeweils ein erstes Bild (In1, In2, In3, ...) mit Regen-, Lichteinfall und/oder Schmutzbeeinträchtigung und ein zweites Bild (Out1, Out2, Out3, ...)
20 ohne Beeinträchtigung zeitgleich oder unmittelbar aufeinander folgend erfasst werden.
3. Verfahren nach Anspruch 1 oder 2, wobei ein künstliches neuronales Netzwerk (CNN1; CNN10, CNN11, CNN12) für alle Kameras (2-i) gemeinsam trainiert wird.
4. Verfahren nach einem der vorhergehenden Ansprüche, wobei die
30 Trainings-Bildpaare (In1, Out1; In2, Out2; In3, Out3; ...) mindestens eine Folge von aufeinander folgenden Eingangs- und Ausgangsbildern enthalten.
5. Verfahren zum Korrigieren von Eingangs-Bilddaten einer Mehrzahl von Kameras (2-i) eines Rundumsichtsystems, die durch Regen, Lichteinfall und/oder Schmutz
35 beeinträchtigt sind, mit den Schritten:

a) von den Kameras (2-i) erfasste Eingangs-Bilddaten (Ini), die durch Regen, Lichteinfall und/oder Schmutz beeinträchtigt sind, werden einem trainierten künstlichen neuronalen Netzwerk (CNN1, CNN10, CNN11, CNN12) bereitgestellt,

b) das trainierte künstliche neuronale Netzwerk (CNN1, CNN10, CNN11, CNN12) ist dazu konfiguriert, die Eingangs-Bilddaten (Ini) mit Regen-, Lichteinfall und/oder Schmutzbeeinträchtigung in korrigierte Ausgabe-Bilddaten (Opti) ohne Beeinträchtigung umzuwandeln und ein Sicherheitsmaß c zu ermitteln, das vom Grad der Benetzung durch Wasser, Lichteinfall und/oder Verschmutzung für ein Bild der Eingangs-Bilddaten abhängig ist und die Sicherheit des Netzwerks charakterisiert, dass die Bildkorrektur des Netzwerks zutreffend ist, und

c) das trainierte künstliche neuronale Netzwerk (CNN1, CNN10, CNN11, CNN12) ist dazu konfiguriert, die Ausgabe-Bilddaten (Opti) und das ermittelte Sicherheitsmaß c auszugeben.

6. Verfahren nach Anspruch 5, wobei die Eingangs-Bilddaten (Ini) mindestens eine Folge von aufeinander folgend erfassten Eingangsbildern als Eingangs-Bilddaten (Ini) enthalten.

7. Verfahren nach einem der Ansprüche 5 oder 6, wobei die Kameras (2-i) fahrzeuggebundene Umfelderfassungskameras sind.

8. Verfahren nach Anspruch 7, wobei eine Ausgabe der umgewandelten Bilddaten (Opti) und des ermittelten Sicherheitsmaßes c an mindestens eine ADAS-Detektionsfunktion erfolgt, welche auf der Grundlage der umgewandelten Bilddaten ADAS-relevante Detektionen ermittelt und ausgibt.

9. Vorrichtung mit mindestens einer Datenverarbeitungseinheit konfiguriert zum Korrigieren von Eingangs-Bilddaten (Ini) einer Mehrzahl von Kameras (2-i) eines Rundumsichtsystems, die durch Regen, Lichteinfall und/oder Schmutz beeinträchtigt sind, in korrigierte Ausgabe-Bilddaten (Opti) umfassend:

- eine Eingangsschnittstelle, die konfiguriert ist zum Empfangen der Eingangs-Bilddaten (Ini) mit Regen-, Lichteinfall und/oder Schmutzbeeinträchtigung von den Kameras (2-i),
- ein trainiertes künstliches neuronales Netzwerk (CNN1, CNN10, CNN11, CNN12), das dazu konfiguriert ist, die Eingangs-Bilddaten (Ini) in korrigierte Ausgabe-Bilddaten (Opti) ohne Beeinträchtigung umzuwandeln und ein Sicherheitsmaß c zu ermitteln, das vom Grad der Benetzung durch Wasser, Lichteinfall und/oder Verschmutzung für ein Bild der

Eingangs-Bilddaten abhängig ist und die Sicherheit des Netzwerks charakterisiert, dass die Bildkorrektur des Netzwerks zutreffend ist, und

- eine erste Ausgangsschnittstelle (CNN11), die dazu konfiguriert ist, die umgewandelten Ausgabe-Bilddaten (Opti) und das ermittelte Sicherheitsmaß c auszugeben.

10. Vorrichtung nach Anspruch 9, wobei die Datenverarbeitungseinheit in eine hardwarebasierter Bildvorverarbeitungsstufe implementiert ist.

11. Vorrichtung nach Anspruch 9 oder 10, wobei die Kameras (2-i) fahrzeuggebundene Umfelderkassungskameras sind und das trainierte künstliche neuronale Netzwerk (CNN1, CNN10, CNN11) zur Bildkorrektur Bestandteil eines fahrzeugseitigen ADAS-Detektions-Neuronalen-Netzwerks (CNN2, CNN12) mit einer geteilten Eingangsschnittstelle, und zwei separaten Ausgangsschnittstellen ist, wobei die erste Ausgangsschnittstelle (CNN11) zur Ausgabe der korrigierten Ausgabebilddaten (Opti) und die zweite Ausgangsschnittstelle (CNN12) zur Ausgabe der ADAS-relevanten Detektionen konfiguriert sind.

12. Vorrichtung nach einem der Ansprüche 9 bis 11, wobei die Eingangsbilddaten mindestens eine Folge von aufeinander folgend erfassten Eingangsbildern als Eingangs-Bilddaten (Ini) enthält und das künstliche neuronale Netzwerk (CNN1, CNN10, CNN11, CNN12) anhand von mindestens einer Folge von aufeinander folgenden Eingangs- und Ausgangsbildern trainiert worden ist.

13. Computer Programmelement, welches, wenn damit eine Datenverarbeitungseinheit programmiert wird, die Datenverarbeitungseinheit dazu anweist, ein Verfahren nach einem der Ansprüche 5 bis 8 durchzuführen.

14. Computerlesbares Speichermedium, auf dem ein Programmelement nach Anspruch 13 gespeichert ist.

15. Verwendung eines Verfahrens nach Anspruch 1 bis 4 zum Trainieren eines künstlichen neuronalen Netzwerks (CNN1, CNN10, CNN11, CNN12) einer Vorrichtung nach einem der Ansprüche 9 bis 12.

1/3

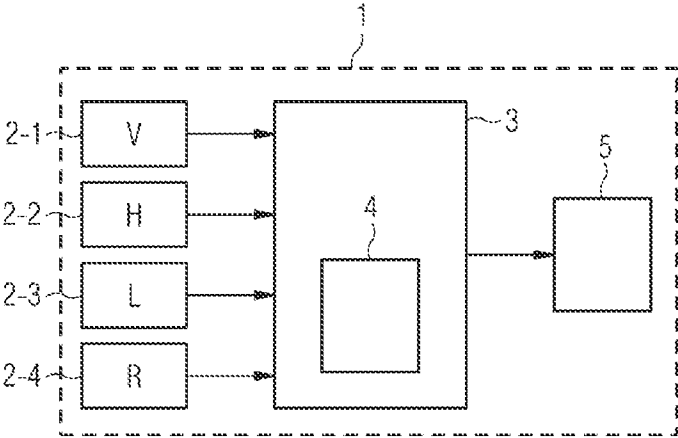


Fig. 1

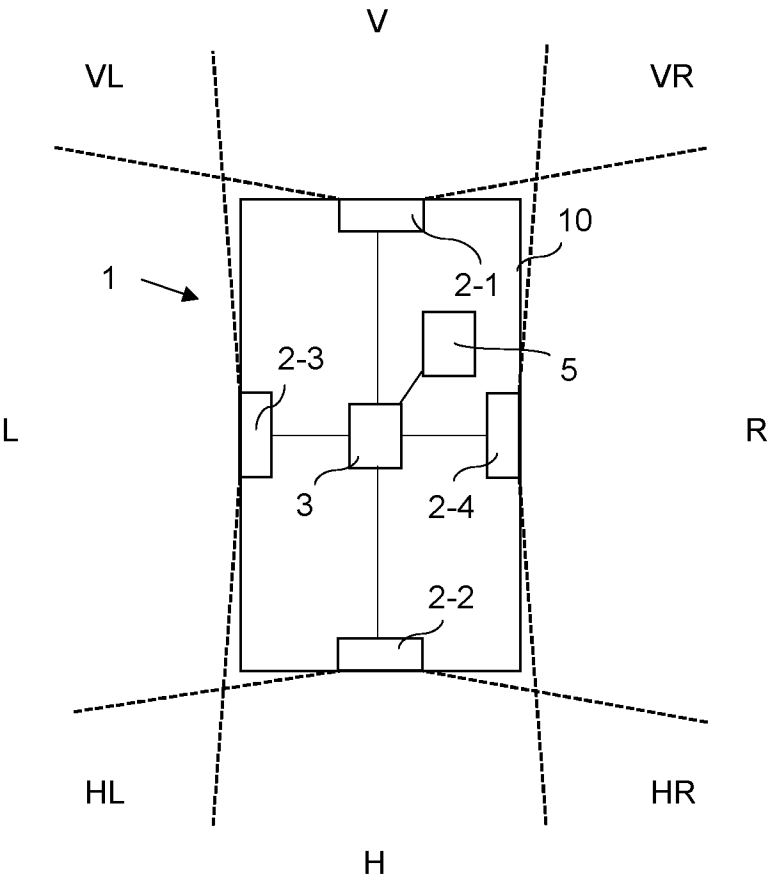


Fig. 2

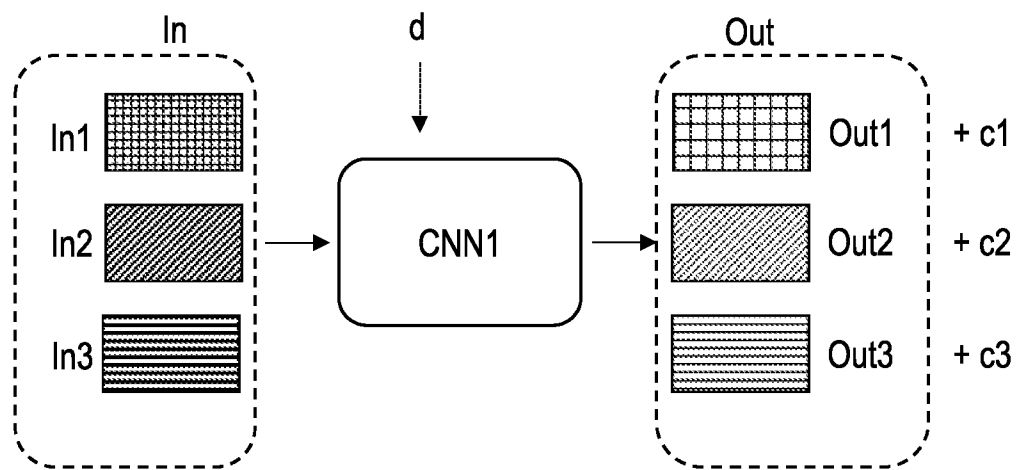
2/3

Fig. 3

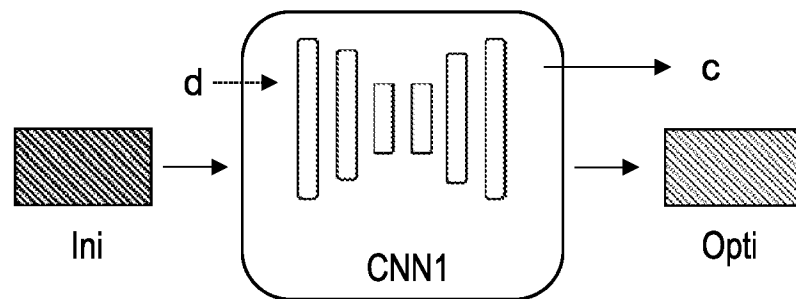


Fig. 4

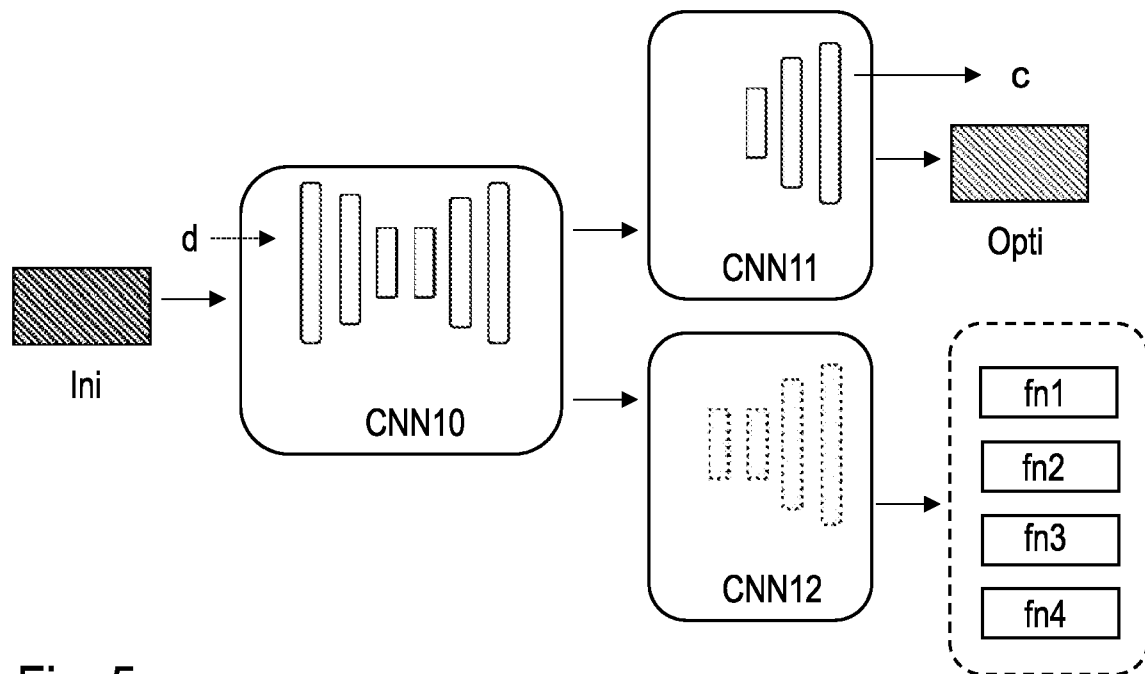
3/3

Fig. 5

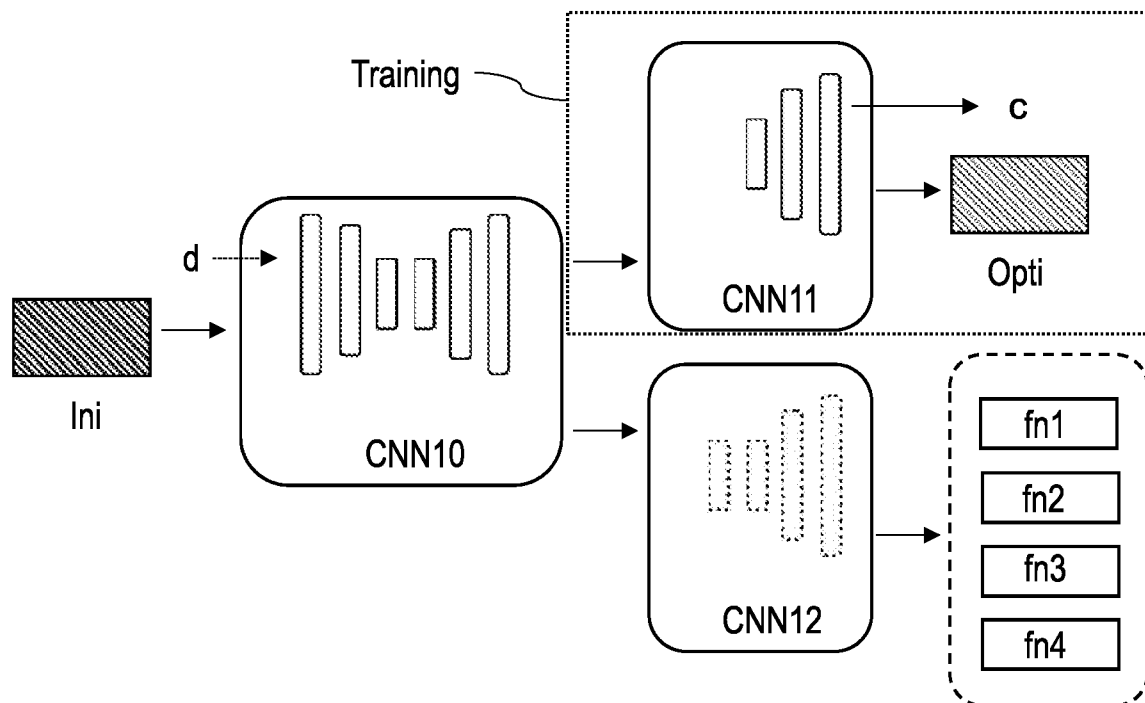


Fig. 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/DE2021/200236

A. CLASSIFICATION OF SUBJECT MATTER**G06V 20/56**(2022.01)i; **G06V 10/82**(2022.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06V

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DE 102019205962 A1 (BOSCH GMBH ROBERT [DE]) 29 October 2020 (2020-10-29) paragraph [0002] paragraph [0005] - paragraph [0006] paragraph [0042] paragraph [0009] paragraph [0027] - paragraph [0028]	1-15
X	ALLETTO STEFANO ET AL. "Adherent Raindrop Removal with Self-Supervised Attention Maps and Spatio-Temporal Generative Adversarial Networks" <i>2019 IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION WORKSHOP (ICCVW)</i> , <i>IEEE</i> , 27 October 2019 (2019-10-27), pages 2329-2338 DOI: 10.1109/ICCVW.2019.00286 XP033732390 abstract; figures 5, 6	1,9



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

08 April 2022

Date of mailing of the international search report

22 April 2022

Name and mailing address of the ISA/EP

European Patent Office
p.b. 5818, Patentlaan 2, 2280 HV Rijswijk
Netherlands

Telephone No. (+31-70)340-2040

Facsimile No. (+31-70)340-3016

Authorized officer

Mitzel, Dennis

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/DE2021/200236

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LIU XING ET AL. "Dual Residual Networks Leveraging the Potential of Paired Operations for Image Restoration" <i>2019 IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), IEEE</i> , 15 June 2019 (2019-06-15), pages 7000-7009 DOI: 10.1109/CVPR.2019.00717 XP033686685 the whole document	1-15
A	HORIA PORAV ET AL. "Rainy screens: Collecting rainy datasets, indoors" <i>ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853</i> , 10 March 2020 (2020-03-10), XP081618496 the whole document	1-15
A	US 2020051217 A1 (SHEN LIYING [US] ET AL) 13 February 2020 (2020-02-13) the whole document	1-15
A	US 2020204732 A1 (YADAV TARUN [IN] ET AL) 25 June 2020 (2020-06-25) the whole document	1-15
A	PORAV HORIA ET AL. "I Can See Clearly Now: Image Restoration via De-Raining" <i>2019 INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION (ICRA), IEEE</i> , 20 May 2019 (2019-05-20), pages 7087-7093 DOI: 10.1109/ICRA.2019.8793486 XP033593449 the whole document	1-15

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/DE2021/200236

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
DE	102019205962	A1	29 October 2020	CN	111860761	A	30 October 2020
				DE	102019205962	A1	29 October 2020
				US	2020342574	A1	29 October 2020
US	2020051217	A1	13 February 2020	CN	112703509	A	23 April 2021
				EP	3834135	A1	16 June 2021
				JP	2021534520	A	09 December 2021
				KR	20210059712	A	25 May 2021
				US	2020051217	A1	13 February 2020
				US	2020051260	A1	13 February 2020
				US	2022044363	A1	10 February 2022
				WO	2020033524	A1	13 February 2020
US	2020204732	A1	25 June 2020	NONE			

A. KLASSTIFIZIERUNG DES ANMELDUNGSGEGENSTANDES

INV. G06V20/56 G06V10/82

ADD.

Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

G06V

Recherchierte, aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, WPI Data**C. ALS WESENTLICH ANGESEHENE UNTERLAGEN**

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	DE 10 2019 205962 A1 (BOSCH GMBH ROBERT [DE]) 29. Oktober 2020 (2020-10-29) Absatz [0002] Absatz [0005] – Absatz [0006] Absatz [0042] Absatz [0009] Absatz [0027] – Absatz [0028] ----- -/-	1-15



Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen



Siehe Anhang Patentfamilie

* Besondere Kategorien von angegebenen Veröffentlichungen :

"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

"E" frühere Anmeldung oder Patent, die bzw. das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

"X" Veröffentlichung von besonderer Bedeutung: die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

"Y" Veröffentlichung von besonderer Bedeutung: die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

8. April 2022

Absendedatum des internationalen Recherchenberichts

22/04/2022

Name und Postanschrift der Internationalen Recherchenbehörde

Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Mitzel, Dennis

C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	ALLETTO STEFANO ET AL: "Adherent Raindrop Removal with Self-Supervised Attention Maps and Spatio-Temporal Generative Adversarial Networks", 2019 IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION WORKSHOP (ICCVW), IEEE, 27. Oktober 2019 (2019-10-27), Seiten 2329-2338, XP033732390, DOI: 10.1109/ICCVW.2019.00286 [gefunden am 2020-03-02] Zusammenfassung; Abbildungen 5, 6 -----	1, 9
A	LIU XING ET AL: "Dual Residual Networks Leveraging the Potential of Paired Operations for Image Restoration", 2019 IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), IEEE, 15. Juni 2019 (2019-06-15), Seiten 7000-7009, XP033686685, DOI: 10.1109/CVPR.2019.00717 [gefunden am 2020-01-08] das ganze Dokument -----	1-15
A	HORIA PORAV ET AL: "Rainy screens: Collecting rainy datasets, indoors", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 10. März 2020 (2020-03-10), XP081618496, das ganze Dokument -----	1-15
A	US 2020/051217 A1 (SHEN LIYING [US] ET AL) 13. Februar 2020 (2020-02-13) das ganze Dokument -----	1-15
A	US 2020/204732 A1 (YADAV TARUN [IN] ET AL) 25. Juni 2020 (2020-06-25) das ganze Dokument -----	1-15
A	PORAV HORIA ET AL: "I Can See Clearly Now: Image Restoration via De-Raining", 2019 INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION (ICRA), IEEE, 20. Mai 2019 (2019-05-20), Seiten 7087-7093, XP033593449, DOI: 10.1109/ICRA.2019.8793486 [gefunden am 2019-08-09] das ganze Dokument -----	1-15

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/DE2021/200236

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
DE 102019205962 A1	29-10-2020	CN 111860761 A	30-10-2020
		DE 102019205962 A1	29-10-2020
		US 2020342574 A1	29-10-2020

US 2020051217 A1	13-02-2020	CN 112703509 A	23-04-2021
		EP 3834135 A1	16-06-2021
		JP 2021534520 A	09-12-2021
		KR 20210059712 A	25-05-2021
		US 2020051217 A1	13-02-2020
		US 2020051260 A1	13-02-2020
		US 2022044363 A1	10-02-2022
		WO 2020033524 A1	13-02-2020

US 2020204732 A1	25-06-2020	KEINE	
