



(10) **DE 10 2021 002 363 B3** 2022.09.08

(12)

Patentschrift

(21) Aktenzeichen: **10 2021 002 363.3**
(22) Anmeldetag: **04.05.2021**
(43) Offenlegungstag: –
(45) Veröffentlichungstag
der Patenterteilung: **08.09.2022**

(51) Int Cl.: **G01N 21/88** (2006.01)
G01B 11/24 (2006.01)
G01B 11/30 (2006.01)

Innerhalb von neun Monaten nach Veröffentlichung der Patenterteilung kann nach § 59 Patentgesetz gegen das Patent Einspruch erhoben werden. Der Einspruch ist schriftlich zu erklären und zu begründen. Innerhalb der Einspruchsfrist ist eine Einspruchsgebühr in Höhe von 200 Euro zu entrichten (§ 6 Patentkostengesetz in Verbindung mit der Anlage zu § 2 Abs. 1 Patentkostengesetz).

(73) Patentinhaber:
**Karlsruher Institut für Technologie, Körperschaft
des öffentlichen Rechts, 76131 Karlsruhe, DE**

(74) Vertreter:
**Müller-Boré & Partner Patentanwälte PartG mbB,
80639 München, DE**

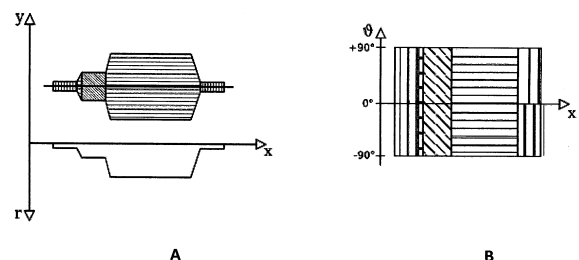
(72) Erfinder:
**Mitschke, Norbert, 67346 Speyer, DE; Heizmann,
Michael, 76344 Eggenstein-Leopoldshafen, DE**

(56) Ermittelter Stand der Technik:

DE	10 2018 103 244	A1
DE	10 2020 111 252	A1
US	2019 / 0 073 568	A1
EP	2 850 386	B1

(54) Bezeichnung: **Gerät für eine lageunabhängige optische Oberflächeninspektion**

(57) Zusammenfassung: Die Erfindung betrifft ein computerimplementiertes Verfahren und eine Vorrichtung zur optischen Oberflächeninspektion von Objekten. Das Verfahren umfasst: Bereitstellen eines statistischen Segmentierungsmodells, welches anhand einer Mehrzahl von Trainingsbildern eines Objekts in unterschiedlichen Zuständen trainiert worden ist, wobei das Segmentierungsmodell als Eingang ein Bild nimmt und als Ergebnis eine Klassifikation der Pixel im Eingangsbild in zumindest einer Klasse liefert, Bestimmen einer mit einem Parameter p parametrisierbaren Trajektorie anhand zumindest eines von einer Kamera aufgenommenen und mit Hilfe des Segmentierungsmodells segmentierten Bilds eines zu inspizierenden Objekts; Aufnahme einer zeitlichen Abfolge von Bildern des zu inspizierenden Objekts von der Kamera während sich die Kamera auf der Trajektorie um das Objekt bewegt; Segmentieren der Bilder der zeitlichen Abfolge mit Hilfe des Segmentierungsmodells; Extrahieren zumindest einer zu inspizierenden (Teil-)Oberfläche des Objekts aus jedem der von der Kamera aufgenommenen Bilder mit Hilfe des Segmentierungsergebnis für das jeweilige Bild; Fusionieren der extrahierten (Teil-)Oberflächen in einer zwei-dimensionalen Darstellung, wobei die zwei-dimensionale Darstellung den Parameter p und eine zur Trajektorie orthogonale Bildkoordinate als Ortskoordinaten aufweist; und Durchführen einer Defekterkennung auf die fusionierte zwei-dimensionale Darstellung.



Beschreibung

[0001] Die Erfindung betrifft eine Vorrichtung und ein Verfahren für eine lageunabhängige optische Oberflächeninspektion von Objekten.

[0002] Insbesondere betrifft die Erfindung eine Vorrichtung für eine lageunabhängige optische Oberflächeninspektion von Objekten (wie z.B. Werkstücken) mittels Einsatzes von tiefen neuronalen Netzwerken auf Videosequenzen.

[0003] Vorrichtungen und Verfahren zur optischen Oberflächeninspektion von Werkstücken sind aus dem Stand der Technik bekannt. In der Veröffentlichung [1] werden Videodaten mittels neuronaler Netzwerke für die Erkennung/ Segmentierung und Inspektion von Brückenbauteilen verwendet. Hierbei werden vergangene Frames verwendet, um die Klassifikation des aktuellen Frames u.a. durch Kontextinformationen zu verbessern. Das Training des neuronalen Netzwerks erfolgt auf simulierten Daten. Als neuronales Netzwerk wird ein RNN und ein FCN verwendet, bei denen die örtliche Auflösung des Eingangsbildes reduziert werden. Eine Defekterkennung der extrahierten Komponenten findet nicht statt.

[0004] In der Veröffentlichung [2] wird ein Verfahren beschrieben, dass die Bilder eines Videos von Brennstoffkanälen in Kernkraftwerken zusammensetzt („Stitching“) und so in der Lage ist, ein 2D-Bild der Innenbohrung zu erzeugen. Dies ermöglicht ein besseres Erkennen von Anomalien im Zeitverlauf. Die Bilder wurden bei bekannter Orientierung der Kamera (0°, 60°, 120°, 180°, 240°, 300°) im Kanal aufgenommen. Die so erhaltenen 6 Streifen werden mithilfe der Kenntnis von Position, Orientierung und Geschwindigkeit der Kamera fusioniert („Knowledge-based image stitching“). In diesem Verfahren werden die Positionen der Kamera für das Zusammensetzen der Bilder vorausgesetzt.

[0005] In der Veröffentlichung [3] wird ein neuronales Netz für Erkennung, Segmentierung und Defekterkennung von Fehlern auf Eisenbahnschienen vorgestellt, das Bilder von einem Gleisbett als Eingabe hat. Es wird aus einer Serie von Bildern jedes Frame einzeln untersucht. Die Informationserfassung findet daher auf Einzelbildern und nicht auf Videosequenzen statt.

[0006] In der Veröffentlichung [4] wird ein Sichtprüfgerät dargestellt, das ein Werkstück, das sich in einer definierten Position befindet, mittels Videoabtastung inspiziert.

[0007] Die Veröffentlichungen [5] bis [7] beschreiben Anwendungen, in denen neuronale Netze für die visuelle Inspektion / Defekterkennung verwendet

werden. Dabei wird aber keine Kamera verwendet, die sich zur Bildaufnahme bewegt.

[1]: Narazaki, Yasutaka, et al. „Automated bridge component recognition using video data.“ arXiv preprint arXiv: 1806.06820 (2018).

[2]: Murray, Paul, et al. „Automated image stitching for enhanced visual inspections of nuclear power stations.“ (2013).

[3]: Gibert, Xavier, Vishal M. Patel, and Rama Chellappa. „Deep multitask learning for railway track inspection.“ IEEE transactions on intelligent transportation systems 18.1 (2016): 153-164.

[4]: Nielson, Paul C., and Wolfgang Kaufman. „Machine visual inspection device and method.“ U.S. Patent No. 4,760,444. 26 Jul. 1988.

[5]: Bastian, Blossom Treasa, et al. „Visual inspection and characterization of external corrosion in pipelines using deep neural network.“ NDT & E International 107 (2019): 102134.

[6]: Li, Bing, and Biao Jiang. „Deep Neural Network based Visual Inspection with 3D Metric Measurement of Concrete Defects using Wall-climbing Robot.“ (2019).

[7]: Nagata, F., et al. „Basic application of deep convolutional neural network to visual inspection.“ Proceedings of International Conference on Industrial Application Engineering (ICIAE2018). 2018.

[0008] Bei den herkömmlichen Verfahren zur optischen Oberflächeninspektion von Objekten (wie z.B. von Werkstücken) werden für die Inspektion meist mehrere Bilder des Objekts einzeln untersucht bzw. Zeilenkameras verwendet. Alternativ wird ein fusioniertes Bild verwendet, das durch gezieltes Abfahren vordefinierter Punkte und anschließendem Scannen des Objektes entsteht. Beide Verfahren haben einen hohen Anspruch an die Kalibrierung und setzen die initiale Kenntnis der Objektposition sowie der zeitabhängigen Lage der Kamera voraus.

[0009] Aufgabe der Erfindung ist es, ein verbessertes Verfahren und eine entsprechende Vorrichtung zur optischen Oberflächeninspektion von Objekten bereitzustellen, die es ermöglichen, eine optische lageunabhängige Inspektion durchzuführen.

[0010] Diese Aufgabe wird durch eine Vorrichtung zur optischen Oberflächeninspektion, ein computerimplementiertes Verfahren zur optischen Oberflächeninspektion und ein Computerprogrammprodukt mit den in den unabhängigen Ansprüchen angegebenen Merkmalen gelöst. Bevorzugte Ausführungsvarianten bzw. -formen sind Gegenstand der abhängigen Ansprüche.

[0011] Gemäß einem ersten Aspekt der Erfindung wird eine Vorrichtung zur optischen Oberflächeninspektion von Objekten bereitgestellt. Die Vorrichtung umfasst:

eine bewegbare Kamera (wie z.B. eine Videokamera), welche auf einer mit einem Parameter p parametrisierbare Trajektorie um ein zu inspizierendes Objekt bewegbar ist, wobei die Kamera eingerichtet ist, während ihrer Bewegung auf der Trajektorie eine zeitliche Abfolge (z.B. in Form eines Videos bzw. einer Videosequenz) von Bildern des Objekts aus unterschiedlichen Positionen auf der Trajektorie aufzunehmen;

eine Rechenvorrichtung umfassend:

eine Modellbereitstellungseinheit, welche eingerichtet ist, ein statistisches Segmentierungsmodell bereitzustellen, wobei das Segmentierungsmodell anhand einer Mehrzahl von Trainingsbildern des Objekts in unterschiedlichen Zuständen trainiert worden ist, wobei das Segmentierungsmodell als Eingang ein Bild nimmt und als Ergebnis eine Klassifikation der Pixel im Eingangsbild in zumindest einer Klasse liefert, wobei die zumindest eine Klasse einer zu inspizierenden (Teil-)Oberfläche des Objekts entspricht.

eine Bildsegmentierungseinheit, die eingerichtet ist, ein von der Kamera aufgenommenes Bild (Eingangsbild für die Segmentierung) mit Hilfe des Segmentierungsmodells zu segmentieren;

eine Trajektorienbestimmungseinheit, die eingerichtet ist, die Trajektorie anhand zumindest eines von der Kamera aufgenommenen und von der Bildsegmentierungseinheit segmentierten Bilds des Objekts zu bestimmen;

eine Bildtransformationseinheit, die eingerichtet ist, aus jedem Bild der zeitlichen Abfolge von Bildern, welche von der Kamera bei ihrer Bewegung auf der Trajektorie aufgenommen worden sind, zumindest eine zu inspizierende (Teil-)Oberfläche des Objekts anhand des Segmentierungsergebnis für das Bild zu extrahieren und die extrahierten (Teil-)Oberflächen in einer zwei-dimensionalen (planaren) Darstellung zu fusionieren, wobei die zwei-dimensionale Darstellung den Parameter p und eine zur Trajektorie orthogonale Bildkoordinate als Ortskoordinaten aufweist;

eine Defekterkennungseinheit, die eingerichtet ist, eine Defekterkennung auf die fusionierte zwei-dimensionale Darstellung durchzuführen. Die Defekterkennung kann mittels eines herkömmlichen Defekterkennungsverfahrens erfolgen.

[0012] Die Vorrichtung zur optischen Oberflächeninspektion von Objekten verwendet ein Segmentierungsmodell, welches sowohl für die Extraktion der zu inspizierenden (Teil-)Oberflächen des Objekts als auch für die Bestimmung einer objekttypischen, parametrisierbaren Trajektorie, auf welcher sich die Kamera bewegt, verwendet wird.

[0013] Die bewegbare Kamera nimmt bei ihrer Bewegung auf der Trajektorie Bilder des Objekts aus unterschiedlichen Richtungen und optional aus unterschiedlichen Entfernungen zwischen Kamera und Objekt auf. Die Position und/oder Orientierung des Objekts zur Kamera brauchen dabei nicht weiter spezifiziert und/oder beschränkt oder vorab vorgegeben zu sein. Diese können anhand der von der Kamera aufgenommenen Bilder mit Hilfe des Segmentierungsmodells ermittelt werden.

[0014] Mit der Vorrichtung zur optischen Oberflächeninspektion gemäß dem ersten Aspekt ist es somit möglich, Unregelmäßigkeiten von im Allgemeinen beliebigen Objekten und insbesondere von im Wesentlichen rotationssymmetrischen Objekten durch Zusammenführen von Einzelbildern ohne Kenntnis über die Position von Objekt und Kamera zuverlässig und schnell zu erkennen.

[0015] Im Unterschied zu bekannten Vorrichtungen und Verfahren zur optischen Oberflächeninspektion von Objekten muss keine genaue Lage der Kamera oder des Objektes bekannt sein. Ferner wird eine zeitliche Abfolge (z.B. in Form eines Videos bzw. einer Videosequenz) und keine Reihe von aus vorbestimmten Positionen aufgenommenen Einzelbildern mit Hilfe des Segmentierungsmodells segmentiert, transformiert und für das Zusammenfügen der zu inspizierenden (Teil-)Oberfläche verwendet. Die Anforderungen an die Kalibrierung der Kamera können folglich erheblich gesenkt werden, was eine schnelle und zuverlässige Inspektion ermöglicht.

[0016] Die Rechenvorrichtung kann ferner eine Modellinferenzeinheit für eine Inferenz bzw. für ein Herleiten des Segmentierungsmodells umfassen. Die Modellinferenzeinheit ist eingerichtet, ein Trainieren eines ursprünglichen, untrainierten Segmentierungsmodells durchzuführen, wobei das Trainieren mit einer Mehrzahl von Trainingsbildern des Objekts in unterschiedlichen Zuständen erfolgt. Ferner kann die Rechenvorrichtung eine Trainingsbildeingabeschnittstelle zum Bereitstellen der Trainingsbilder umfassen.

[0017] Das Segmentierungsmodell bzw. die Bildsegmentierungseinheit nimmt als Eingang ein Bild und liefert als Ergebnis eine Klassifikation der Pixel im Eingangsbild in zumindest einer Klasse, wobei die zumindest eine Klasse einer zu inspizierenden (Teil-)Oberfläche des Objekts entspricht. Anders ausge-

drückt liefert das Segmentierungsmodell als Ausgang eine modellierte (z.B. vorhergesagte oder geschätzte) Klassenzugehörigkeit für die Pixel des Eingangsbilds. Bei der Segmentierung eines Bilds mit Hilfe des Segmentierungsmodells enthält jedes Bildpixel die Information, zu welcher Klasse es gehört. Dabei können mehrere Klassen definiert werden. Die Klassen können zum Beispiel mehreren zu inspizierenden (Teil-)Oberfläche(n) des Objekts entsprechen (je eine Klasse für eine zu inspizierenden (Teil-)Oberfläche). Umfasst das Objekt mehrere Komponenten, kann das Segmentierungsmodell eine Klassifizierung der Pixel des Eingangsbilds zu mehreren Klassen liefern, wobei jede Klasse einer bestimmten Komponente entspricht. Zusätzlich kann eine „Dummy“-Klasse definiert werden, in der alle Pixel klassifiziert werden, die zu keiner bestimmten Komponente zugeordnet werden können. Die Summe der Pixel, welche in den einzelnen Klassen (z.B. in den einzelnen Komponenten-Klassen und gegebenenfalls in der „Dummy“- Klasse) ergibt dann das Segmentierungsergebnis für das Objekt als Ganzes. Es ist auch möglich, für das Objekt eine eigene, separate Klassifizierungsklasse zu definieren.

[0018] Die modellierte Klassenzugehörigkeit für jedes Pixel des Eingangsbilds kann z.B. in Form von Segmentierungsmasken ausgegeben werden. Eine Segmentierungsmaske entspricht der Gruppe von Pixeln, welche einer bestimmten Klasse (wie z.B. einem bestimmten Objekt, einer bestimmten Komponente und/oder einer bestimmten zu inspizierenden (Teil-)Oberfläche des Objekts) zugeordnet werden. Im Falle eines Ankers als Objekt können die unterschiedlichen (Teil-)Oberflächen die Kommutatoroberfläche, die Wellenoberfläche und/oder die Ritzeloberfläche sein. Allgemein im Fall eines im Wesentlichen rotationssymmetrischen Objekts, kann die zu inspizierende Oberfläche des Objekts die Mantelfläche des Objekts oder einen Teil der Manteloberfläche sein.

[0019] Das Segmentierungsmodell kann ein statistisches Modell sein, das anhand einer Vielzahl von Trainingsbildern trainiert worden ist. Das Segmentierungsmodell kann zum Beispiel mehrere Modellparameter (wie z.B. die Gewichtungen eines neuronalen Netzwerks) umfassen, welche während des Trainings mit den Trainingsbildern verändert bzw. angepasst werden. Beispielhafte statistische Modelle sind lineare und nicht lineare Regressionsmodelle (wie zum Beispiel lineare Regression, nichtlineare Regression, nichtlineare Regression mit einem Attention-Mechanismus, nichtlineare multi-Task Regression, nicht-parametrische oder semiparametrische Regression, etc.), Klassifikationsmodelle, Modelle des maschinellen Lernens, etc.

[0020] Beispielsweise kann das Segmentierungsmodell aus einem neuronalen Netz bestehen oder ein neuronales Netz umfassen, wobei das neuronale Netz mit den Trainingsbildern des Objekts und die zur Inspektion relevanten (Teil-)Oberflächen des Objekts trainiert worden ist. Beispielhafte neuronale Netze sind tiefe neuronale Netze, CNN („Convolutional Neural Network“ oder faltendes neuronales Netz), U-Netz, etc. Ein geeignetes neuronales Netzwerk zur Segmentierung von Eingangsbildern von Objekten (in dem konkreten Fall von Ankern von Elektromotoren) ist in den Publikationen [8] und [9] beschrieben, deren entsprechende Ausführungen insoweit einen integralen Offenbarungsbestandteil der vorliegenden Anmeldung darstellen:

[8] N. Mitschke und M. Heizmann, „Semantische Segmentierung von Ankerkomponenten von Elektromotoren“, 2020. Forum Bildverarbeitung 2020. Ed.: T. Längle ; M. Heizmann, 329-340, KIT Scientific Publishing;

[9] N. Mitschke und M. Heizmann, „Image-Based Visual Servoing of Rotationally Invariant Objects Using a U-Net Prediction“, 2021. 7th International Conference on Automation, Robotics and Applications (ICARA), 235-240, Institute of Electrical and Electronics Engineers (IEEE).doi:10.1109/ICARA51699.2021.9376577.

[0021] In einem Beispiel kann das Segmentierungsmodell (z.B. das neuronale Netz) auf einer niedrigeren Auflösung arbeiten bzw. mit niedrig aufgelösten Trainingsbildern des Objekts in unterschiedlichen Zuständen trainiert worden sein. Das Segmentierungsergebnis kann ferner mittels klassischer Bildverarbeitung und/oder durch die Verwendung eines semantischen Objektmodells verbessert bzw. stabilisiert werden. Bei der Inspektion eines bestimmten Objekts können Bilder des Objekts mit einer höheren Auflösung von der Kamera aufgenommen und die relevanten Oberflächen des zu untersuchenden Objekts mit Hilfe des Segmentierungsmodells (wie z.B. des trainierten neuronalen Netzwerks) extrahiert werden.

[0022] Das (trainierte) Segmentierungsmodell kann zum Beispiel in einer Speichervorrichtung gespeichert sein, wie zum Beispiel in einer Datenbank, einem Rechner und/oder einer Daten- oder Rechnerwolke.

[0023] Die Vorrichtung kann ferner eine Kamerabewegungseinrichtung umfassen, die eingerichtet ist, die Kamera zu bewegen, und in Signalverbindung mit der Trajektorienbestimmungseinheit steht. Die Kamerabewegungseinrichtung kann zum Beispiel einen Roboterarm umfassen. Ferner kann die Vorrichtung eine Plattform umfassen, auf der das Objekt

liegt und/oder eine Objektbewegungseinheit (z.B. einen Roboterarm).

[0024] Wenn das Eingangsbild für die Segmentierung eine Auflösung aufweist, welche höher ist als die Auflösung der Segmentierung bzw. des Segmentierungsmodells, kann das Eingangsbild zunächst auf die Auflösung der Segmentierung umgetastet werden. Anschließend kann das Segmentierungsmodell auf das umgetastete Bild angewandt werden.

[0025] Wie oben beschrieben wird die Trajektorie anhand zumindest eines von der Kamera aufgenommenen und von der Bildsegmentierungseinheit segmentierten Bilds des Objekts bzw. anhand des Ergebnisses aus der Segmentierung zumindest eines von der Kamera aufgenommenen Bild bestimmt.

[0026] In einem Beispiel kann die Trajektorie derart bestimmt werden, dass alle auf der Trajektorie liegenden Punkte einen vorbestimmten oder vorbestimmbaren Abstand zu einer Achse (wie z.B. einer langen Achse) des Objekts aufweisen.

[0027] Um die Achse zu bestimmen, kann die Trajektorienbestimmungseinheit eine Hauptkomponentenanalyse des segmentierten Bilds durchführen, wobei die Achse anhand des Ergebnisses der Hauptkomponentenanalyse bestimmt wird. Somit ist es möglich, automatisch und effizient eine objektspezifische Trajektorie zu bestimmen, auf der sich die Kamera bewegt und Bilder für die Inspektion des Objekts aufnimmt.

[0028] In einem Beispiel können alle Punkte auf der Trajektorie einen konstanten Abstand zur Achse aufweisen. In diesem Fall ist die Trajektorie ein Kreis mit der ermittelten Achse des Objekts als Normale durch den Mittelpunkt des Kreises. Die Trajektorie kann jedoch eine Kurve sein, bei der der Abstand zur Achse variiert bzw. nicht konstant ist. Der Abstand kann fest vorgegeben oder veränderbar bzw. anpassbar sein. Der Abstand kann somit als ein Freiheitsgrad zur Bestimmung der Trajektorie dienen.

[0029] Vor der Durchführung der Hauptkomponentenanalyse kann das segmentierte Bild einer weiteren Bildtransformation unterzogen werden, z.B. um die Kanten zu verbessern. Dabei können unterschiedliche Kantenverbesserungsverfahren eingesetzt werden. Um die Kanten zu verbessern, kann zum Beispiel das Segmentierungsergebnis mit einer Tiefenkarte des Objekts kombiniert werden. Eine Tiefenkarte des Objekts wird beispielsweise bei der Verwendung einer Tiefenkamera automatisch erzeugt. Alternativ kann die Rechenvorrichtung eingerichtet sein, eine Tiefenkarte des Objekts anhand einer Mehrzahl der von der Kamera aufgenommenen Bilder zu erzeugen. In der Regel sind die Kanten der

Objekte, die anhand der Tiefenkarte ermittelt werden, genauer als die Kanten aus der Segmentierung. Dies erleichtert und erhöht die Genauigkeit der Achsenbestimmung.

[0030] Die Rechenvorrichtung kann beispielsweise eingerichtet sein, eine Tiefenkarte des Objekts bereitzustellen; aus der Tiefenkarte die Tiefenkanten aller Objekte, die sich von einer Bodenebene erheben, zu bestimmen, und die Kanten des Objekts mit dem größten Schnitt mit dem Objekt aus der Segmentierung durch die ermittelten Tiefenkanten zu verbessern.

[0031] Mit Hilfe des Segmentierungsmodells ist es ferner möglich, eine optimale Orientierung und/oder Entfernung der Kamera zum Objekt und/oder Anfangsposition der Kamera zum Objekt zu bestimmen. Die Vorrichtung kann dementsprechend eine Kameraorientierungseinheit umfassen, welche eingerichtet ist, eine (optimale) Orientierung und/oder Entfernung der Kamera zum Objekt und/oder eine (optimale) Anfangsposition der Kamera zum Objekt anhand zumindest eines von der Kamera aufgenommenen und von der Bildsegmentierungseinheit segmentierten Bilds des Objekts zu bestimmen. Die Anfangsposition kann auf der ermittelten Trajektorie liegen.

[0032] Eine optimale Orientierung der Kamera in Bezug auf das Objekt kann z.B. eine Orientierung sein, bei der die optische Achse der Kamera im Wesentlichen orthogonal zu einer (z.B. langen) Achse des Objektes steht. Eine optimale Entfernung zwischen der Kamera und dem Objekt kann z.B. eine Entfernung sein, bei welcher alle relevanten zu inspizierenden (Teil-)Oberflächen im Bild sichtbar sind und eine Defektdetektion ermöglichen. Die optimale Entfernung kann von der Auflösung und/oder dem Öffnungswinkel der Kamera abhängen.

[0033] Für eine optimale Orientierung und/oder Entfernung der Kamera zum Objekt können beispielsweise die Erfüllung einer oder mehrerer der folgenden Bedingungen aufgestellt werden:

- Das Objekt befindet sich in einer vorbestimmten Position im Bild, z.B. im Zentrum des Bilds;
- Das Objekt nimmt einen vorbestimmten Anteil im Bild ein, z.B. 20%, 30%, 50%, etc.;
- Das Objekt weist eine vorbestimmte Orientierung im Bild auf. Zum Beispiel kann gefordert werden, dass eine (lange) Achse des Objekts im Wesentlichen parallel zu einer der Bildkoordinaten ist;

[0034] Insbesondere können aus einem mit Hilfe des Segmentierungsmodells segmentierten Bild des Objekts Lageparameter der Kamera zum Objekt

bestimmt werden. Die Lageparameter umfassen die Position und Orientierung des Objektes relativ zur Kamera und dem Koordinatensystem, das von dieser definiert wird. Anhand der Lageparameter kann die Kamera entsprechend ausgerichtet und zu einer entsprechenden Position zum Objekt bewegt werden. Dies kann z.B. mittels eines bildbasierten Reglers erfolgen. Ein beispielhafter bildbasierter Regler ist z.B. in der Publikation [9] beschrieben.

[0035] Bei der Bewegung der Kamera auf der so ermittelten Trajektorie kann die Kamera automatisch die ermittelte (optimale) Orientierung zum Objekt einnehmen.

[0036] Da zwischen den Einzelbildern der von der Kamera aufgenommenen zeitlichen Abfolge von Bildern eine Bewegung der Kamera stattfindet, ergeben sich teilüberlappende Bilder derselben Szene, welche die zumindest eine zu inspizierende (Teil-)Oberfläche des Objekts umfasst.

[0037] Die Bildtransformationseinheit kann eingerichtet sein, vor dem Extrahieren eine Transformation jedes der von der Kamera aufgenommenen Bilder und der entsprechenden segmentierten Bilder durchzuführen.

[0038] Die Transformation kann zum Beispiel eine oder mehrere der folgenden Operationen umfassen:

Entrollen des jeweiligen Bilds, wobei das Entrollen ein Projizieren der zumindest einen zu inspizierenden (Teil-)Oberfläche des Objekts auf eine planare Ebene umfasst;

Kompensieren der perspektivischen Verzerrung des jeweiligen Bilds; und

Normalisieren des jeweiligen Bilds.

[0039] Die Transformation kann zum Beispiel dazu dienen, die perspektivische Verzerrung der aus unterschiedlichen Positionen der Kamera aufgenommenen Bilder herauszurechnen bzw. zu kompensieren. Ferner ist die zu inspizierende (Teil-)Oberfläche im Allgemeinen eine gekrümmte Fläche. Die Transformation kann folglich dazu dienen, die gekrümmte (Teil-)Oberfläche in einer planaren Fläche zu transformieren. Beim Entrollen wird folglich die Krümmung der zu inspizierenden (Teil-)Oberfläche des Objekts bzw. die gekrümmte Oberflächengeometrie herausgerechnet, um eine planare Fläche zu erhalten. In einer beispielhaften Ausführung werden nur die Pixel im jeweiligen Bild entrollt bzw. transformiert, welche zu der zu inspizierenden (Teil-)Oberfläche oder dem zu inspizierenden Objekt gehören.

[0040] Grundsätzlich werden beide Bilder, d.h. das von der Kamera aufgenommene Bild und das entsprechende segmentierte Bild transformiert. Die Kompensation der perspektivischen Verzerrung der

aus unterschiedlichen Positionen der Kamera aufgenommenen Bilder kann vor dem Entrollen, d.h. vor dem Projizieren der zumindest einen zu inspizierenden (Teil-)Oberfläche des Objekts auf eine planare Fläche bzw. Ebene erfolgen.

[0041] Das Entrollen des von der Kamera aufgenommenen und optional normalisierten Bilds und des entsprechenden segmentierten und optional normalisierten Bilds, kann eine Transformation bzw. eine Neuordnung zumindest einer der Koordinaten (z.B. der y-Koordinate) des jeweiligen Bilds umfassen, um die zumindest eine zu inspizierenden (Teil-)Oberfläche des Objekts auf die planare bzw. plane Ebene zu projizieren. Insbesondere kann das Entrollen des von der Kamera aufgenommenen Bilds und des entsprechenden segmentierten Bilds die folgenden Operationen umfassen:

Bestimmen aus dem segmentierten Bild einer (z.B. langen) Achse des Objekts mittels Hauptkomponentenanalyse;

Bestimmen zumindest eines Abstands von der Achse zu einer Kante des Objekts im segmentierten Bild;

Bestimmen einer Koordinatentransformation (z.B. zumindest einer der beiden Bildkoordinaten, wie z.B. der y-Koordinate) anhand des zumindest einen Abstands;

Transformieren des von der Kamera aufgenommenen Bilds und des entsprechenden segmentierten Bilds anhand der Koordinatentransformation.

[0042] Das Normalisieren kann beispielsweise eine oder mehrere der folgenden Operationen umfassen:

- Normalisieren der Orientierung des Objekts im Bild, z.B. durch eine Schätzung der Orientierung des Objekts im Bild und eine Rotation, sodass die Achse des Objekts im Bild parallel zur einer der Koordinatenachsen (wie z.B. der x-Achse) des Bilds ist;

- Normalisieren der Position und/oder des Anteils des zu untersuchenden Objekts im Bild, z.B. durch Zuschneiden („Crop“-Operation),

- Normalisieren der Größe des Bilds oder des Bildausschnitts, sodass das Bild oder der Bildausschnitt eine vorgegebene, feste Größe (z.B. in Pixeln) aufweist (z.B. durch eine „Resize“-Operation).

[0043] Vorzugsweise erfolgt das Normalisieren vor dem Entrollen des Bilds.

[0044] Die oben beschriebenen Transformationen und insbesondere das Entrollen der Bilder vor der Extraktion erleichtern und verbessern die Extraktion der relevanten zu inspizierenden (Teil-)Oberflächen

aus den von der Kamera aufgenommenen Bildern sowie das Fusionieren der extrahierten (Teil-)Oberflächen in einer zwei-dimensionalen (planaren) Darstellung, die nachfolgend auf Defekte untersucht wird.

[0045] Um die aus unterschiedlichen Positionen entnommenen Bilder der zumindest einen zu inspizierenden (Teil-)Oberfläche des Objekts zu fusionieren, können unterschiedliche bekannte Bildregistrierungs- und Fusionierungsverfahren eingesetzt werden.

[0046] Das Fusionieren der extrahierten Bilder in der zwei-dimensionalen Darstellung kann zum Beispiel die folgenden Operationen umfassen:

für jedes Paar von zwei aufeinanderfolgenden Bildern, die von der Kamera aufgenommen und optional transformiert worden sind:

Bestimmen von einer Mehrzahl von Merkmalen an bestimmten Punkten in jedem der Bilder;

Auffinden von Punktepaaren mit übereinstimmenden Merkmalen (z.B. mittels einer Merkmalsanpassung („Feature-Matching“) in den zwei aufeinanderfolgenden Bildern;

Bestimmen einer (z.B. affinen) Transformationsmatrix zwischen den aufeinander folgenden Bildern;

Fusionieren der aufeinanderfolgenden Bilder, indem eines der Bilder mit Hilfe der Transformationsmatrix in Deckung zum anderen Bild gebracht wird, wobei im Bereich der Überschneidung der Teil des jeweiligen Bildes genommen wird, der näher am Zentrum des jeweiligen Bilds ist.

[0047] Die obigen Schritte werden für alle Bilderpaare der zeitlichen Abfolge von Bildern wiederholt.

[0048] Alternativ kann das Fusionieren der extrahierten Bilder in der zwei-dimensionalen Darstellung die folgenden Operationen umfassen:

für jedes Paar von zwei aufeinanderfolgenden Bildern, die von der Kamera aufgenommen und optional transformiert worden sind:

Auffinden einer Translation eines der Bilder gegenüber dem anderen Bild, die eine Korrelationsfunktion zwischen den beiden Bildern maximiert (hierzu kann Template-Matching verwendet werden); und

Fusionieren der aufeinanderfolgenden Bilder, indem eines der Bilder mit Hilfe der Translation in Deckung zum anderen Bild gebracht wird, wobei im Bereich der Überschneidung der Teil des jeweiligen Bildes genommen wird, der näher am Zentrum des jeweiligen Bilds ist.

[0049] Die obigen Schritte werden für alle Bilderpaare der zeitlichen Abfolge von Bildern wiederholt.

[0050] Gemäß einem zweiten Aspekt der Erfindung wird ein computerimplementiertes Verfahren zur optischen Oberflächeninspektion von Objekten bereitgestellt. Das Verfahren kann zum Beispiel mit Hilfe der oben beschriebenen Vorrichtung durchgeführt werden. Das Verfahren umfasst:

Bereitstellen eines statistischen Segmentierungsmodells, welches anhand einer Mehrzahl von Trainingsbildern eines Objekts in unterschiedlichen Zuständen trainiert worden ist, wobei das Segmentierungsmodell als Eingang ein Bild nimmt und als Ergebnis eine Klassifikation der Pixel im Eingangsbild in zumindest einer Klasse liefert, wobei die zumindest eine Klasse einer zu inspizierenden (Teil-)Oberfläche des Objekts entspricht.

Bestimmen einer mit einem Parameter p parametrisierbaren Trajektorie um ein zu inspizierendes Objekt anhand zumindest eines von einer Kamera aufgenommenen und mit Hilfe des Segmentierungsmodells segmentierten Bilds des Objekts;

Aufnahme einer zeitlichen Abfolge von Bildern des zu inspizierenden Objekts von einer Kamera, welche sich auf der Trajektorie um das Objekt bewegt und während ihrer Bewegung Bilder des Objekts aus unterschiedlichen Positionen auf der Trajektorie aufnimmt;

Segmentieren der Bilder der zeitlichen Abfolge von Bildern des Objekts mit Hilfe des Segmentierungsmodells;

Extrahieren zumindest einer zu inspizierenden (Teil-)Oberfläche des Objekts aus jedem der von der Kamera aufgenommenen Bilder mit Hilfe des Segmentierungsergebnisses für das jeweilige Bild;

Fusionieren der extrahierten (Teil-)Oberflächen in einer zwei-dimensionalen (planaren) Darstellung, wobei die zwei-dimensionale Darstellung den Parameter p und eine zur Trajektorie orthogonale Bildkoordinate als Ortskoordinaten aufweist;

Durchführen einer Defekterkennung auf die fusionierte zwei-dimensionale Darstellung.

[0051] Das Bestimmen der Trajektorie kann die folgenden Operationen umfassen:

Bestimmen einer (z.B. langen) Achse des Objekts mit Hilfe einer Hauptkomponentenanalyse des segmentierten Bilds, und

Bestimmen der Trajektorie derart, dass die auf der Trajektorie liegenden Punkte einen vorbe-

stimmten oder vorbestimmbaren Abstand zur ermittelten Achse aufweisen.

[0052] Ferner kann das Verfahren umfassen:

Bereitstellen einer Tiefenkarte des Objekts

Bestimmen von Tiefenkanten aller Objekte, die sich von einer Bodenebene erheben anhand der Tiefenkarte, und

Verbessern (z.B. Verstärken, Begradigen) der Kanten des Objekts mit dem größten Schnitt mit dem Objekt aus der Segmentierung mit Hilfe der ermittelten Tiefenkanten.

[0053] Des Weiteren kann das Verfahren umfassen:

Bestimmen einer (optimalen) Orientierung der Kamera zum Objekt anhand zumindest eines von der Kamera aufgenommenen und mit Hilfe des Segmentierungsmodells segmentierten Bilds des Objekts.

[0054] Des Weiteren kann das Verfahren Transformieren jedes der von der Kamera aufgenommenen Bilder und der entsprechenden segmentierten Bilder vor dem Extrahieren umfassen, wobei das Transformieren eine oder mehrere der folgenden Operationen umfasst:

Entrollen des jeweiligen Bilds umfassend Projizieren der zumindest einen zu inspizierenden (Teil-)Oberfläche des Objekts auf eine planare Ebene;

Kompensieren der perspektivischen Verzerrung des jeweiligen Bilds; und

Normalisieren des jeweiligen Bilds.

[0055] Das Entrollen des von der Kamera aufgenommenen Bilds und des entsprechenden segmentierten Bilds kann die folgenden Operationen umfassen:

Bestimmen aus dem segmentierten Bild einer (z.B. langen) Achse des Objekts mittels Hauptkomponentenanalyse;

Bestimmen zumindest eines Abstands von der Achse zu einer Kante des Objekts im segmentierten Bild;

Bestimmen einer Koordinatentransformation (z.B. zumindest einer der beiden einen Bildkoordinaten, wie z.B. der y-Koordinate) anhand des zumindest einen Abstands; und

Transformieren des von der Kamera aufgenommenen Bilds und des entsprechenden segmentierten Bilds anhand der Koordinatentransformation.

[0056] Ein beispielhaftes Fusionieren der extrahierten Bilder in der zwei-dimensionalen Darstellung kann die folgenden Operationen umfassen:

für jedes Paar von zwei aufeinanderfolgenden Bildern, die von der Kamera aufgenommen und optional transformiert worden sind:

Bestimmen von einer Mehrzahl von Merkmalen an bestimmten Punkten in jedem der Bilder;

Auffinden von Punktepaaren mit übereinstimmenden Merkmalen (z.B. mittels Merkmalsanpassung („feature matching“) in den zwei aufeinanderfolgenden Bildern;

Bestimmen einer (z.B. affinen) Transformationsmatrix zwischen den aufeinander folgenden Bildern;

Fusionieren der aufeinanderfolgenden Bilder, indem eines der Bilder mit Hilfe der Transformationsmatrix in Deckung zum anderen Bild gebracht wird, wobei im Bereich der Überschneidung der Teil des jeweiligen Bildes genommen wird, der näher am Zentrum des jeweiligen Bilds ist.

[0057] Die obigen Schritte werden für alle Bilderpaare der zeitlichen Abfolge von Bildern wiederholt.

[0058] Ein anderes beispielhaftes Fusionieren der extrahierten Bilder in der zwei-dimensionalen Darstellung kann die folgenden Operationen umfassen:

für jedes Paar von zwei aufeinanderfolgenden Bildern, die von der Kamera aufgenommen und optional transformiert worden sind:

Auffinden einer Translation eines der Bilder gegenüber dem anderen Bild, die eine Korrelationsfunktion zwischen den beiden Bildern maximiert (hierzu kann Template-Matching verwendet werden); und

Fusionieren der aufeinanderfolgenden Bilder, indem eines der Bilder mit Hilfe der Translation in Deckung zum anderen Bild gebracht wird, wobei im Bereich der Überschneidung der Teil des jeweiligen Bildes genommen wird, der näher am Zentrum des jeweiligen Bilds ist.

[0059] Die obigen Schritte werden für alle Bilderpaare der zeitlichen Abfolge von Bildern wiederholt.

[0060] In Bezug auf das Verfahren gemäß dem zweiten Aspekt und der oben beschriebenen Beispiele gelten sinngemäß die in Zusammenhang mit der oben beschriebenen Vorrichtung aufgeführten Definitionen, Ausführungsformen und Varianten, Beispiele und Vorteile.

[0061] Gemäß einem dritten Aspekt wird ein Computerprogrammprodukt bzw. Computerprogramm

bereitgestellt, welches eingerichtet ist, wenn geladen und ausgeführt auf einem Computer ein beispielhaftes Verfahren zur optischen Oberflächeninspektion von Objekten gemäß dem zweiten Aspekt und der oben beschriebenen Beispiele durchzuführen.

[0062] Das Computerprogrammprodukt bzw. Computerprogramm kann auf einem körperlichen Speichermedium bzw. Programmträger gespeichert werden. Das Computerprogrammprodukt kann ferner als Programmsignal vorliegen.

[0063] In Bezug auf das Computerprogrammprodukt bzw. Computerprogramm gelten sinngemäß die in Zusammenhang mit der oben beschriebenen Vorrichtung aufgeführten Definitionen, Ausführungsformen und Varianten, Beispiele und Vorteile.

[0064] Die oben genannten Vorrichtungen (wie z.B. die Rechenvorrichtung) und Einheiten zum Bereitstellen, Bestimmen, Festlegen oder Berechnen von Daten (wie z.B. Modellbereitstellungseinheit, Trajektorienbestimmungseinheit, Bildsegmentierungseinheit, Bildtransformationseinheit, Defekterkennungseinheit etc.) können durch geeignet konfigurierte bzw. programmierte Datenverarbeitungsvorrichtungen (insbesondere spezialisierte Hardwaremodule, Computer oder Computersysteme, wie zum Beispiel Rechner- oder Datenwolken) mit entsprechenden Recheneinheiten, elektronische Schnittstellen, Speicher und Datenübermittlungseinheiten realisiert werden.

[0065] Die oben genannten Vorrichtungen und/oder Einheiten können in Signalverbindung miteinander stehen und Daten austauschen (wie z.B. die Ergebnisse der Segmentierung, Trajektorienbestimmung, Bildtransformationen etc.). Ferner können die oben genannten Vorrichtungen und/oder Einheiten geeignete Schnittstellen aufweisen, die eine Übermittlung, Eingabe und/oder Auslesen von Daten (wie zum Beispiel Trainingsbildern, Kamerabildern, Parametern der Trajektorie etc.) ermöglichen. Ebenfalls können die Vorrichtungen zumindest eine Speichereinheit umfassen, zum Beispiel in Form einer Datenbank, welche die verwendeten Daten speichert.

[0066] Die Vorrichtungen und/oder Einheiten können ferner zumindest eine vorzugsweise interaktive grafische Benutzerschnittstelle (GUI) umfassen, welche es einem Benutzer ermöglicht, Daten zu betrachten und/oder einzugeben und/oder zu modifizieren.

[0067] Nachfolgend werden bevorzugte Ausführungsformen der vorliegenden Erfindung anhand begleitender Figuren beispielhaft beschrieben. Einzelelemente der beschriebenen Ausführungsformen sind nicht auf die jeweilige Ausführungsform beschränkt. Vielmehr können Elemente der Ausführungsformen beliebig miteinander kombiniert werden

und neue Ausführungsformen dadurch erstellt werden. Es zeigen:

Fig. 1A ein beispielhaftes Trainingsbild eines Objekts;

Fig. 1B eine beispielhafte Annotierung des in **Fig. 1** gezeigten Trainingsbilds;

Fig. 2A eine schematische Ansicht eines beispielhaften Objekts;

Fig. 2B eine schematische Ansicht einer beispielhaften Anordnung Kamera-Objekt;

Fig. 3A eine schematische Ansicht einer beispielhaften optimale Lage der Kamera zum Objekt;

Fig. 3B ein beispielhaftes Bild des Objekts, das von der Kamera in der optimalen Lage aufgenommen worden ist;

Fig. 4A ein beispielhaftes Bild eines zu inspizierenden Objekts;

Fig. 4B eine beispielhafte planare Darstellung, welche anhand einer zeitlichen Abfolge von Bildern des Objekts ermittelt worden ist;

Fig. 5A die Maske eines Ankers, die aus einem beispielhaften segmentierten Bild abgeleitet werden kann ;

Fig. 5B das in **Fig. 5A** gezeigte Bild mit den mittels einer Hauptkomponentenanalyse ermittelten Eigenvektoren;

Fig. 5C das in **Fig. 5B** gezeigte und normalisierte Bild;

Fig. 6 ein schematisches Abbild eines beispielhaften Objekts 10;

Fig. 7A ein von der Kamera aufgenommene Bild eines beispielhaften Objekts;

Fig. 7B das Ergebnis der Segmentierung des in **Fig. 7A** gezeigten Bilds des Objekts;

Fig. 8A das Ergebnis der Normalisierung des in **Fig. 7A** gezeigten Bilds;

Fig. 8B das Ergebnis der Normalisierung des in **Fig. 7B** gezeigten segmentierten Bilds;

Fig. 8C schematisch die Bestimmung einer Achse anhand des in **Fig. 8B** gezeigten segmentierten Bilds;

Fig. 9A das Ergebnis des Entrollens des in **Fig. 8A** gezeigten normalisierten Bilds;

Fig. 9B das Ergebnis des Entrollens des in **Fig. 8B** gezeigten segmentierten und normalisierten Bilds;

Fig. 10A eine schematische Ansicht eines Bilds aus der zeitlichen Abfolge von Bildern eines beispielhaften Objekts;

Fig. 10B das Ergebnis der Fusion der aus der zeitlichen Abfolge von Bildern extrahierten (Teil-)Oberfläche des in **Fig. 10** gezeigten Objekts;

Fig. 11 ein beispielhaftes Verfahren zum Registrieren von Kamerabildern.

[0068] Nachfolgend wird ein beispielhaftes Verfahren zur lageunabhängigen optischen Oberflächeninspektion eines Objekts (z.B. eines Werkstücks) beschrieben. Das Verfahren kann auf zwei verschiedenen Auflösungsstufen arbeiten.

[0069] In einer ersten Phase (Modellerstellungsphase oder Lernphase) wird zunächst anhand von Bildern des Objekts in unterschiedlichen Zuständen ein Segmentierungsmodell des Objekts erstellt. Das Ergebnis (d.h. der Ausgang des Segmentierungsmodells) ist zumindest eine Segmentierungsmaske des Objekts und/oder der zu inspizierenden (Teil-)Oberflächen des Objekts. Die Bilder, mit denen das Segmentierungsmodell erstellt wird, werden im Rahmen der vorliegenden Anmeldung Trainingsbilder genannt.

[0070] Die unterschiedlichen Zustände des Objekts können beispielsweise unterschiedliche Bauformen des Objekts, unterschiedliche Positionen zwischen dem Objekt und der Kamera etc. sein. Die Auflösung der Trainingsbilder kann relativ niedrig sein, z.B. 112x112 bis 448x448 Pixel. Die Anzahl der Trainingsbilder kann z.B. 25 bis 200 sein. Eine höhere oder niedrigere Anzahl ist ebenfalls möglich.

[0071] In einer zweiten Phase (Inspektionsphase oder Untersuchungsphase) wird ein bestimmtes Objekt auf Defekte untersucht. Hierzu wird eine zeitliche Abfolge von Bildern des Objekts (z.B. in Form eines Videos) von einer beweglichen Kamera aufgenommen. Die aufgenommenen Bilder können eine zweite Auflösung aufweisen, welche höher als die erste Auflösung ist und eine Defektdetektion ermöglicht. Beispielsweise kann die erste Auflösung zwischen 112x112 bis 448x448 Pixel liegen und muss nicht zwangsläufig mit der Auflösung der Trainingsbilder übereinstimmen und die zweite Auflösung kann beispielsweise im Bereich 1333x750 bis 3266x1837 Pixel liegen. Aus den von der Kamera aufgenommenen Bildern des zu untersuchenden Objekts werden mit Hilfe des Segmentierungsmodells das Objekt und die zu inspizierenden (Teil-)Oberflächen des Objekts extrahiert. Die extrahierten Oberflächen werden registriert und fusioniert, sodass sie auf Defekte untersucht werden können.

[0072] Das Segmentierungsmodell kann ein statistisches Modell sein, das anhand der Trainingsbilder trainiert wird. Beispielsweise kann das Segmentierungsmodell ein trainiertes neuronales Netz sein oder ein solches Netz umfassen, wobei das neuro-

nale Netz auf die Segmentierung von Bildern des Objekts in unterschiedlichen Zuständen, vorzugsweise mit Augmentierung, trainiert worden ist. Die für die Inspektion relevanten (Teil-)Oberflächen des Objekts können separat in der Lernphase gelernt werden. Im Beispiel eines Ankers können dies die Welle und die Kommutatorfläche sein.

[0073] Um die Parameter und die Segmentierungsmasken zu stabilisieren, kann ferner ein semantisches Modell bei der Segmentierung verwendet werden. Dabei können frühere Segmentierungsergebnisse und/oder Plausibilitäten (z.B.: ein Anker kann maximal 1 Kommutator haben) verwendet werden. Hierfür werden die Masken durch geeignete Merkmalsextraktoren (z.B. ORB-Features) zur Übereinstimmung gebracht werden (Kontrollpunktanpassung oder „keypoint matching“). Des Weiteren können die Segmentierungsmasken weiterer Bildbearbeitungsverfahren unterzogen werden, um z.B. die Kanten zu verbessern.

[0074] **Fig. 1** zeigt eine beispielhafte Annotierung eines Objekts (eines Ankers) beim Erstellen des Segmentierungsmodells, wobei **Fig. 1A** das Originalbild und **Fig. 1B** das mit Hilfe eines Segmentierungsmodells annotierte Bild des Objekts zeigt. Bei diesem Beispiel werden in der Lernphase die folgenden Komponenten und die entsprechenden (Teil-)Oberflächen gelernt: Anker, Welle, Kommutator und Ritzel. Das verwendete Segmentierungsmodell ist ein neuronales Netz, wie z.B. in der Publikation [8] beschrieben.

[0075] Wenn ein bestimmtes Objekt untersucht werden soll, wird eine zeitliche Abfolge (z.B. in Form eines Videos) von vorzugsweise hochauflösten Bildern des Objekts von einer beweglichen Kamera aufgenommen. Die Kamera bewegt sich auf einer objekttypischen Trajektorie und nimmt eine zeitliche Abfolge von Bildern des Objekts aus unterschiedlichen Positionen auf der Trajektorie auf. **Fig. 2A** zeigt eine schematische seitliche Ansicht eines beispielhaften Objekts 10 (eines Ankers) und **Fig. 2B** eine beispielhafte Anordnung Kamera-Objekt bei der Bewegung der Kamera 12 auf einer objekttypischen Trajektorie 14.

[0076] Um ein gutes Bild für die Extraktion und die Registrierung zu erhalten, kann zunächst eine geeignete (optimale) Position und/oder Orientierung der Kamera 12 in Bezug auf das Objekt 10 bestimmt werden. Dies kann zum Beispiel mittels bildbasierter Regelung erfolgen. Hierzu kann aus dem Segmentierungsergebnis zumindest eines mit der Kamera aufgenommenen Bildes des zu untersuchenden Objekts die Parameter der Position und/oder Orientierung der Kamera in Bezug auf das Objekt bestimmt (z.B. geschätzt) werden. Diese Parameter

werden im Rahmen der vorliegenden Anmeldung auch Lageparameter genannt. Beispielhafte Lageparameter sind Größe, Orientierung und Neigung zwischen der Objektebene und der Kameraebene. Ein beispielhaftes Verfahren zur bildbasierten Regelung ist in der Publikation [8] beschrieben.

[0077] Eine beispielhafte optimale Position und Orientierung der Kamera 12 zum Objekt ist in **Fig. 3A** gezeigt. **Fig. 3B** zeigt ein beispielhaftes Bild des Objekts, das von der Kamera 12 in der optimalen Position und Orientierung zum Objekt 10 aufgenommen worden ist. Die Kamera 12 kann mittels eines Roboterarms 16 zu der zuvor bestimmten Position bewegt und/oder entsprechend der zuvor bestimmten Orientierung ausgerichtet werden. Alternativ oder zusätzlich kann auch das zu untersuchende Objekt 10 relativ zur Kamera 12 bewegt werden.

[0078] Befindet sich das Objekt 10 in der zuvor bestimmten Position und/oder Orientierung bezüglich der Kamera 12, wird die Kamera 12 in einer objekttypischen mit dem Parameter p parametrisierbaren Trajektorie 14 über das Objekt 10 bewegt. Alternativ oder zusätzlich kann das Objekt 10 in Bezug auf die Kamera 12 bewegt werden. Das Objekt 10 kann beispielsweise durch einen zweiten Roboterarm gedreht werden, um die Rückseite zu begutachten.

[0079] Für das Beispiel eines rotationssymmetrischen Objekts mit einer (z.B. langen) Achse x , wie z.B. des in **Fig. 2A** gezeigten Objekts, ist eine geeignete Trajektorie ein Teilkreis (wie z.B. ein Halbkreis) oder ein Kreis in der yz -Ebene. Die yz -Ebene ist die Ebene, welche senkrecht zu der Achse x des Objekts steht, wie z.B. in **Fig. 2B** gezeigt. Der Halbkreis oder der Kreis ist durch den Winkel ϑ parametrisierbar. Der Winkel ϑ ist der Winkel zwischen der z -Achse und der Gerade, welche durch den Mittelpunkt des Halbkreises und die jeweilige Position der Kamera auf Trajektorie durchgeht.

[0080] Während sich die Kamera 12 auf der Trajektorie 14 bewegt, werden die von der Kamera aufgenommenen hochauflösten Bilder auf die Auflösung der Segmentierung umgetastet, durch das Segmentierungsmodell segmentiert und optional nachbearbeitet. Das Nacharbeiten kann zum Beispiel das Anwenden eines semantischen Modells und/oder Bildbearbeitung umfassen.

[0081] Mithilfe des Ergebnisses der Segmentierung und optional der Lageparameter kann aus jedem der von der Kamera 12 aufgenommenen hochauflösten Bilder das Objekt extrahiert werden. Die aus den einzelnen Bildern extrahierten Objekte werden anschließend registriert und fusioniert und in ein Koordinatensystem überführt, das durch den Parameter p (wie z.B. den Winkel ϑ) und die zur Trajekto-

rie orthogonale Bildkoordinate aufgespannt wird. Eine Verschiebung der Kameraposition um „Op“ entlang der Trajektorie bzw. eine Änderung des Winkels ϑ führt somit zu einer Translation auf der Ordinatenachse im neuen Koordinatensystem. Das Registrieren und Fusionieren kann z.B. mittels Schablonen- bzw. Mustervergleichs („Template Matching“) erfolgen.

[0082] Die zur Inspektion relevanten (Teil-)Oberflächen werden anschließend aus dem fusionierten Bild des Objekts mit Hilfe des Segmentierungsmodells ausgeschnitten bzw. extrahiert, um diese inspizieren zu können.

[0083] **Fig. 4A** zeigt ein Bild eines beispielhaften Objekts mit überlagerter Segmentierungsmasken. **Fig. 4B** zeigt eine extrahierte, registrierte und zusammengefügte (Teil-)Oberfläche des Objekts in einer zwei-dimensionalen (planaren) Darstellung.

[0084] Bei der Defektdetektion können unterschiedliche Verfahren verwendet werden. Beispielhaft kann eine sogenannte „Novelty Detection“ (z.B. mittels eines neuronalen Netzes) verwendet werden, bei der eine Reihe von Gut-Bildern auf die oben beschriebene Art extrahiert werden und mit denen ein Autoencoder trainiert wird bzw. trainiert worden ist. Fehler (oder sog. „Novelties“) werden durch eine schlechte Performance des Autoencoders erkannt. Klassische Bildverarbeitungsmethoden (mit Hilfe geeignet definierter Merkmale) sind für diese Defektdetektion ebenso möglich.

[0085] Wie oben beschrieben muss im Unterschied zu bekannten Vorrichtungen und Verfahren zur optischen Oberflächeninspektion von Objekten keine genaue Lage der Kamera bzw. des Objektes bekannt sein. Ferner wird eine zeitliche Abfolge (z.B. in Form eines Videos) und keine Reihe von aus vorbestimmten Positionen aufgenommenen Einzelbildern mit Hilfe eines Segmentierungsmodells segmentiert, transformiert und für das Zusammenfügen der zu inspizierenden (Teil-)Oberfläche verwendet.

[0086] Der aktuelle Parameter $p(t_i)$ der Trajektorie muss ferner nicht bekannt sein, da das Zusammenführen nicht auf Vorwissen basiert, sondern z.B. über Template-Matching vollzogen werden kann.

[0087] Die bildbasierte Regelung und die Extraktion können in Echtzeit erfolgen. In diesem Fall wird das Verfahren vorzugsweise auf einer geeignet konfigurierten Recheneinheit mit hoher Parallelisierbarkeit durchgeführt werden, wie z.B. GPUs, TPUs oder entsprechende FPGAs und ASICs. Darüber hinaus kann sich die Kamera flüssig über die Oberfläche bewegen, ohne dass dabei bestimmte Punkte angefahren werden, an denen die Kamera kurz verweilt.

[0088] Nachfolgend werden die einzelnen Schritte eines beispielhaften Verfahrens zur optischen Oberflächeninspektion (wie z.B. das oben beschriebene Verfahren) im Detail beschrieben.

Segmentierungsmodell

[0089] Wie oben beschrieben kann das Segmentierungsmodell ein statistisches Modell sein, welches mit einer Vielzahl von Trainingsbildern trainiert wird. Ein beispielhaftes Segmentierungsmodell kann ein neuronales Netzwerk sein, z.B. ein CNN oder ein anderes tiefes neuronales Netzwerk.

[0090] Ein geeignetes Segmentierungsmodell ist in der Publikation 9 beschrieben. Das Segmentierungsmodell umfasst ein U-Netz, welches als Eingabe ein RGB-Bild nimmt und als Ausgabe die geschätzte Klassenzugehörigkeit für jeden Pixel im Bild liefert. Die namensgebende Form des U-Netzes entsteht durch die tiefer werdenden Merkmalskarten zur Mitte hin und die Querverbindungen, bei denen Merkmalskarten gleicher Größe konkateniert werden. Ein beispielhaftes U-Net hat z.B. eine Eingabegröße von 224×224 Pixeln, fünf Tiefenstufen und ca. 31 Mio. trainierbare Parameter. Auf jede 3 × 3-Faltungsschicht folgt Batch-Normalisierung und eine ReLU-Aktivierung. Beim Hochtasten wird das 2×2 -Interpolationsfilter auch im Training gelernt. Als Zielfunktion wird der generalisierte Sørensen-Dice-Koeffizient verwendet. Dieser bildet die gewichtete Summe der Sørensen-Dice-Koeffizienten oder einzelnen Klassen.

[0091] Um das U-Net zu trainieren, kann eine Vielzahl (z.B. 100, 150, 200, etc.) von Bildern des Objekts verwendet werden. Die Bilder können auch Bilder mit einer relativ niedrigen Auflösung sein. Die Trainingsbilder können vor der Eingabe in das Segmentierungsmodell unterschiedlichen Bildbearbeitungsoperationen unterzogen werden, um die Genauigkeit der Segmentierung zu verbessern. Die Bildbearbeitungsoperationen können eine oder mehrere der folgenden Operationen umfassen: Augmentierung, Regularisierung und Normalisierung.

[0092] Die Augmentierung kann eine oder mehrere der folgenden Augmentierungsoperationen umfassen: Spiegelung, affine Transformationen, Farbmanipulationen, Rauschen, Skalieren, Zuschneiden. Durch die Augmentierung wird verhindert, dass das Segmentierungsmodell auf das „falsche“ Objekt (z.B. den Hintergrund) trainiert wird.

[0093] Ferner können zusätzliche Informationen und/oder semantischen Modelle verwendet werden, um die Genauigkeit der Segmentierung zu erhöhen. Beispielsweise kann die Hinzunahme der Kanteninformation in einer der hinteren Schichten eines tiefen neuronalen Netzwerks mit mehreren Schichten zu

einer Verbesserung führen, da die Objektgrenzen des Objekts im Bild mit den Kanten im Bild zusammenfallen. Zur Kantenextraktion können unterschiedliche bekannte Verfahren eingesetzt werden. Beispielsweise kann der Marr-Hildreth-Operator verwendet werden. Das Ergebnis kann anschließend normiert werden. Die Kante kann zum Beispiel nach der obersten Konkatenierungsschicht entweder hinzuaddiert oder angehängt werden.

[0094] Um den Ausgang des Segmentierungsmodells (z.B. des neuronalen Netzes) zu stabilisieren, kann ein semantisches Modell des Objekts verwendet werden. Dabei können frühere Segmentierungsergebnisse und/oder Plausibilitäten (z.B.: ein Anker kann maximal 1 Kommutator haben) verwendet werden.

[0095] Ein beispielhaftes Verfahren zur Verbesserung des Segmentierungsergebnisses mit Hilfe eines Objektmodells wird in der Publikation [9] beschrieben. Das Verfahren basiert auf der Annahme, dass innerhalb einer Bildabfolge das gleiche Objekt im Bild zu sehen ist. Des Weiteren wird angenommen, dass das Objekt in den nacheinander folgenden Bildern ähnliche Merkmale aufweist.

[0096] In der Regel handelt es sich bei den zu untersuchenden Objekten um zusammenhängende Objekte. Daher kann ferner eine Regularisierung sinnvoll sein, welche lange Konturen bestraft. Somit können Löcher, kleine Fehldetektionen oder andere Artefakte reduziert werden.

Bestimmung einer optimalen Position und/oder Orientierung der Kamera zum Objekt

[0097] Für eine optimale Position und Orientierung der Kamera 12 zum Objekt 10 können eine oder mehrere der folgenden Eigenschaften gefordert werden:

a) Das Objekt 10 soll im Bild vollständig sichtbar sein. Dies impliziert, dass die Kamera 12 derart orientiert wird, dass sie in Richtung des Objektes 10 schaut und dass eine geeignete Entfernung zwischen Kamera 12 und Objekt 10 gewählt wird. Die Orientierung der Kamera 12 kann so gewählt werden, dass das Objekt 10 näherungsweise in der Mitte des Bildes ist.

b) Das Objekt 10 soll im Bild so groß sein, dass relevante Details (z.B. Defekte) erkennbar sind, wobei vorzugsweise das Objekt 10 gleichzeitig vollständig im Bild ist. Diese Anforderung lässt einen gewissen Spielraum bei der Entfernung zwischen Objekt 10 und Kamera 12, der im Wesentlichen durch die Auflösung und den Öffnungswinkel der Kamera bestimmt wird. Um die Anforderungen einzuhalten kann beispielsweise gefordert werden, dass das Objekt 10 einen

konstanten Anteil des Bildes (z.B. 12%, 20%, 30%, 50%, etc.) einnimmt.

c) Die zu untersuchende (Teil-)Oberfläche oder die zu untersuchenden (Teil-)Oberflächen des Objekts 10 soll bzw. sollen sichtbar sein. Bei einem rotationssymmetrischen Objekt 10 kann zum Beispiel gefordert werden, dass die Orientierung zwischen Kamera 12 und Objekt 10 so gewählt ist, dass die Mantelfläche des Objekts 10 sichtbar ist, und dass optional die Grundflächen (die nicht inspiziert werden sollen) nicht sichtbar sind.

[0098] Die Orientierung des Objektes 10 im Bild hingegen kann frei gewählt werden und stellt einen Freiheitsgrad dar. Ohne Beschränkung der Allgemeinheit kann zum Beispiel gefordert werden, dass eine (lange) Achse des Objekts 10 parallel zur x-Achse des Koordinatensystems des Bilds steht.

[0099] Eine beispielhafte optimale Position und Orientierung der Kamera 12 zum Objekt 10 ist in **Fig. 3A** gezeigt. Das von der Kamera in dieser Position aufgenommene Bild ist in **Fig. 3B** gezeigt.

[0100] Wie oben beschrieben, können die optimale Position und Orientierung der Kamera 12 zum Objekt 10 mittels eines bildbasierten Reglers bestimmt werden. Alternativ oder zusätzlich kann die Position und/oder Orientierung der Kamera manuell (entweder durch eine bewegliche Kamera oder durch eine geeignete Positionierung des Objektes bei fester Kamera) oder durch eine Steuerung mittels eines neuronalen Netzes (z.B. mittels „Reinforcement Learning“) bestimmt werden.

Bestimmung der Trajektorie der Kamera

[0101] Die Trajektorie kann aus zumindest einem von der Kamera aufgenommenen Bild des Objekts bestimmt werden. Das Bild, anhand dessen die Trajektorie bestimmt wird, kann ein Bild sein, das aus der oben beschriebenen optimalen Position und Orientierung der Kamera zum Objekt aufgenommen worden ist (siehe z.B. **Fig. 3B**).

[0102] Das Objekt kann zumindest eine zu inspizierende (Teil-)Oberfläche aufweisen. Neben der zumindest einen zu inspizierenden Teiloberfläche kann das Objekt noch weitere Flächen aufweisen, die nicht inspiziert zu werden brauchen.

[0103] Die Trajektorie der Kamera wird derart bestimmt, dass eine Inspektion der (Teil-)Oberfläche bzw. (Teil-)Oberflächen eines Objektes ermöglicht wird. Die Orientierung der Kamera bei ihrer Bewegung auf der Trajektorie kann die oben beschriebene optimale Orientierung sein.

[0104] In einem Beispiel ist das Objekt im Wesentlichen rotationssymmetrisch. Beispielhafte Objekte sind u.a. Zylinder, Kegel und Komposita davon. Abweichungen von der rotationsymmetrischen Form (wie z.B. von der Zylinder- oder Kegelform). Das Verfahren kann auch auf (leicht) gekrümmte, konvexe und (näherungsweise) rotationssymmetrische Objekte übertragen werden, solange die Sichtbarkeit aller relevanten Teile bzw. (Teil-)Oberflächen des Objektes auf einen Großteil oder im Wesentlichen allen Bildern der Trajektorie gegeben ist. Das Objekt kann beispielsweise bis zu einem gewissen Grad S-förmig oder wellenförmig gekrümmt sein,

[0105] Bei einem im Wesentlichen rotationssymmetrischen (z.B. zylinder- oder kegelförmigen) Objekt kann die zu inspizierende Oberfläche die Mantelfläche des Objekts oder ein Teil der Manteloberfläche sein. Die Mantelfläche kann deutlich größer als die Grundflächen sein. Die zwei Grundflächen stellen in diesem Beispiel Flächen dar, die nicht inspiziert werden. Die Grundflächen müssen nicht identisch sein.

[0106] Das Objekt kann eine (z.B. lange) Achse (Rotationsachse), die zum Beispiel näherungsweise mittig durch das Objekt verläuft, aufweisen. Die objekttypische Trajektorie ergibt sich dann als ein Kreis bzw. ein Kreisbogen (z.B. ein Kreisbogen eines Halbkreises) mit der Achse als Normale durch den Mittelpunkt bzw. dem Schwerpunkt des dazugehörigen Kreises und der aktuellen Kameraposition als Punkt auf diesem Kreis bzw. Kreisbogen. Die Achse kann wie nachfolgend beschrieben aus einem mit Hilfe des Segmentierungsmodells segmentierten Bild mittels Hauptkomponentenanalyse geschätzt bzw. bestimmt werden.

Bestimmung der Achse

[0107] Zumindest ein von der Kamera 10 aufgenommenes Bild wird mit Hilfe des Segmentierungsmodells segmentiert. Aus dem segmentierten Bild bzw. des Segmentierungsergebnisses kann die Achse 20 mittels Hauptkomponentenanalyse bestimmt bzw. geschätzt werden.

[0108] Das segmentierte Bild liefert eine Maske 18 mit zum Objekt gehörenden Pixeln i mit den Koordinaten $x_i = [x_{1,i}, x_{2,i}]$ (siehe **Fig. 5A**). Mit dieser Pixelmenge kann eine Hauptkomponentenanalyse durchgeführt werden: Der Bildmittelpunkt ergibt sich aus dem Mittelwert aller Koordinaten der Pixelmenge $\bar{x} = \sum_i x_i$. Mithilfe des Mittelpunktes kann dann eine 2×2 Kovarianzmatrix mit $\text{Cov}(x) = E_i ((x_i - \bar{x})(x_i - \bar{x})^T)$ bestimmt werden. Die Kovarianzmatrix hat zwei Eigenwerte und Eigenvektoren. Der Eigenvektor mit dem größeren Eigenwert korrespondiert zur Richtung, in die das Objekt sich am stärksten ausdehnt. Der andere Eigenwert ist orthogonal zum ersten. Die lange Achse wird dann als Gerade geschätzt, die

durch den Mittelpunkt $\bar{x}$ verläuft und die Richtung des Eigenvektors mit dem größeren Eigenwert hat. Beide Eigenvektoren sind im mittleren segmentierten Bild (siehe **Fig. 5B**) eingezeichnet. Das in **Fig. 5C** gezeigte Bild wurde normalisiert, wobei die Normalisierung eine Rotation des Bildes umfasst, so dass die Achse 20 parallel zu einer der Bildachsen (z.B. der x-Achse) ist. Anhand der mittels Hauptkomponentenanalyse ermittelten (langen) Achse kann die (lange) Achse des Objekts im Raum bzw. in Bezug auf die Kamera bestimmt werden.

[0109] Weitere optionale Schritte der Normalisierung sind das Ausschneiden des relevanten Bildbereichs („Bounding Box“) und die Skalierung des ausgeschnittenen Bildbereichs mittels eines Resize-Befehls, um die Größe (z.B. die absolute Anzahl der Pixel) der Maske konstant zu halten (z.B. 4.000 Pixel).

[0110] Vor der Hauptkomponentenanalyse kann optional das Ergebnis der Segmentierung mithilfe einer Tiefenkarte verbessert werden, da die Kanten der Tiefenkarte i.d.R. genauer sind als bei der Segmentierung. Wird eine Tiefenbildkamera (z.B. Kinect / Zivid) verwendet, wird die Tiefenkarte automatisch von der Kamera geliefert. Aus der Tiefenkarte werden alle Kontouren der Objekte extrahiert, die sich von der Bodenebene abheben. Die Kanten des Objekts mit dem größten Schnitt mit dem Objekt aus der Segmentierung werden durch die Tiefenkartenkanten verbessert.

[0111] Die Bodenebene kann wie folgt geschätzt werden:

Ziel ist es, eine Steigung entlang der x- und eine entlang der y-Achse im Bild zu schätzen. Für jede Zeile im Bild wird zunächst eine Gerade mithilfe eines Least-Square-Schätzers geschätzt. Ausreißer werden hierbei mit RANSAC entfernt. Als Ausreißer können zum Beispiel Fehler in der Tiefenkarte und/oder die Objekte, die auf einer Platte liegen, angenommen werden. Es wird implizit angenommen, dass ein Großteil der Bodenplatte sichtbar ist. Der Mittelwert der Steigung der geschätzten Geraden ist die Steigung m_x in x-Richtung. Mithilfe des y-Achsenabschnitts der geschätzten Geraden und RANSAC wird eine Gerade g_y in y-Achse geschätzt. Die Bodenebene wird durch den y-Achsenabschnitt von g_y , der Steigung g_y und m_x bestimmt.

[0112] Die oben beschriebene Kantenverbesserung des segmentierten Bilds kann auch vor den weiteren Bildtransformationsoperationen (wie z.B. Extrahieren der zumindest einen zu inspizierenden (Teil-)Oberfläche)) durchgeführt werden.

Bewegung der Kamera entlang der Trajektorie

[0113] Die Kamera wird auf der berechneten Trajektorie um das Objekt bewegt und nimmt dabei eine zeitliche Abfolge von Bildern des Objekts auf, die wie beschrieben weiterverarbeitet werden. Die Bilder können kontinuierlich, z.B. in Form eines Videos aufgenommen werden. Dabei ist es nicht erforderlich, dass die Kamera bestimmte vorbestimmte Punkte entlang der Trajektorie anfährt und dort verweilt, um Bilder aufzunehmen.

[0114] Die Trajektorie kann im Wesentlichen konstant gehalten werden. Wird beispielsweise der Abstand zwischen dem Objekt und der Kamera im Wesentlichen konstant gehalten, muss, um auf den nächsten Punkt der Trajektorie zu gelangen, die Kamera einen Schritt auf dem Kreis um das Objekt bewegt werden.

[0115] Es ist möglich, die Trajektorie dynamisch zu verändern. Gemäß der oben beschriebenen optimalen Orientierung der Kamera zum Objekt gibt es drei Freiheitsgrade: die Rotation der Kamera (ändert nur die Orientierung des Objektes im Bild, nicht jedoch die Trajektorie der Kamera), die Elevation zwischen Y- und Z-Achse ϑ (diese bestimmt im Wesentlichen die Trajektorie, siehe z.B. **Fig. 2B** und **Fig. 3A**) und der Abstand Kamera-Objekt. Durch letzteren lässt sich die Trajektorie verändern. Eine Änderung der Trajektorie kann dann sinnvoll sein, wenn Randbedingungen an einem Roboterarm oder einer anderen Kamerabewegungseinrichtung eingehalten werden sollen. Beispielsweise kann die Trajektorie dann verändert werden, wenn ein bestimmter Bereich oder bestimmte Bereiche im Bewegungsbereich des Roboterarms aufgrund von Singularitäten oder drohenden Kollisionen vermieden werden soll bzw. sollen.

Transformation der aufgenommenen Bilder

[0116] Die von der Kamera aufgenommenen Bilder werden optional normalisiert und anschließend entrollt. Bei dem Entrollen der Bilder wird z.B. die y-Achse im Bild transformiert bzw. neu zugeordnet, um eine Schätzung für die planare (Teil-)Oberfläche des Objekts, deren Ausschnitt im Originalbild enthalten ist, zu erhalten.

[0117] In einem Beispiel werden sowohl die von der Kamera aufgenommenen und gegebenenfalls normalisierten Bilder des Objekts als auch die entsprechenden Segmentierungsbilder entrollt. Für die weitere Verarbeitung (wie z.B. Fusionieren) werden jedoch die entrollten Kamerabilder verwendet. Die entsprechenden entrollten segmentierten Bilder dienen zur Prüfung der Plausibilität und können auch weggelassen werden. Es ist ferner nicht notwendig, das gesamte Bild zu entrollen. Es ist ausreichend,

nur die Pixel im Bild, die zum Objekt gehören, zu entrollen, wie bei dem nachfolgenden Beispiel.

[0118] Die Annahme für das Entrollen eines Bilds ist, dass das Bild einen Teil der (Teil-)Oberfläche des Objekts (wie z.B. der Mantelfläche eines im Wesentlichen rotationssymmetrischen Objekts) zeigt, welcher abhängig von der Entfernung zwischen dem Objekt und der Kamera sowie vom Abstand bzw. Radius r von der Achse zur Kante des Objekts im Bild ist.

[0119] Um ein Bild zu entrollen, wird zunächst der Abstand $r(x)$ (z.B. in Pixeln) im Bild geschätzt. Der Abstand $r(x)$ bezeichnet den Abstand von der Achse des Objekts zur Kante des Objekts im Bild, wobei $r_1(x)$ den gemessenen Abstand „nach oben“ (d.h. für $y > 0$), $r_2(x)$ den gemessenen Abstand „nach unten“ (d.h. für $y < 0$), und „ x “ eine Bildkoordinate und „ y “ die zur x -Koordinate orthogonale Bildkoordinate bezeichnet. Der Abstand $r(x)$ bzw. die Abstände $r_1(x)$ und $r_2(x)$ ist bzw. sind im Allgemeinen eine Funktion der x -Koordinate im Bild. Anders ausgedrückt ist der Abstand zwischen der Achse und den Kanten des Objekts im Bild im Allgemeinen nicht konstant bezüglich der Position auf der Achse (x -Koordinate) und bzgl. der Richtung im Bild „nach unten“/ „nach oben“.

[0120] Fig. 6 zeigt ein schematisches Abbild eines beispielhaften Objekts 10 mit einer langen Achse 20 und die Abstände $r_1(x_i)$ und $r_2(x_i)$, wobei x_i die i -ten Pixel entlang der x -Achse bezeichnet.

[0121] Die maximale Größe der im Bild zu sehenden (Teil-)Oberfläche (z.B. Mantelfläche) des Objekts kann bis zu 50% der gesamten (Teil-)Oberfläche (wie z.B. der Mantelfläche) bzw. bis zu 180° betragen. Mithilfe der geschätzten Abstände r_1 und r_2 (z.B. in Pixeln) und des Abstandes der Kamera zum Objekt kann berechnet werden, welcher Anteil der (Teil-)Oberfläche des Objekts (wie z.B. der Mantelfläche des Objekts) im Bild zu sehen ist.

[0122] In einem Beispiel wird angenommen, dass die Entfernung so groß ist, dass der Anteil der zur inspizierenden (Teil-)Oberfläche (wie z.B. der Mantelfläche) im Bild φ° ($\varphi^\circ/2$ nach oben und $\varphi^\circ/2$ nach unten) entspricht. Die y -Koordinate jeder Spalte im Bild (d.h. für ein festes x) ist folgendermaßen verzerrt:

$$\vartheta(x) = \arcsin(y(x)/r_1(x) * \sin(\varphi^\circ/2)) \text{ für } y > 0$$

und

$$\vartheta'(x) = \arcsin(y(x)/r_2(x) * \sin(\varphi^\circ/2)) \text{ für } y < 0,$$

wobei ϑ die neue Koordinate ist.

[0123] Beispielsweise kann $\varphi^\circ = 178^\circ$ sein (89° nach oben und 89° nach unten). In diesem Fall gilt:

$$\vartheta(x) = \arcsin(y(x)/r_1(x) * \sin(89^\circ)) \text{ für } y > 0$$

und

$$\vartheta'(x) = \arcsin(y(x)/r_2(x) * \sin(89^\circ)) \text{ für } y < 0,$$

[0124] Mit Hilfe der obigen Transformation können die von der Kamera aufgenommenen und gegebenenfalls normalisierten Bilder und optional die Masken bzw. die segmentierten Bilder transformiert werden, um entsprechend entrollte Bilder zu erhalten.

[0125] Es ist möglich, die obige Transformation zu verallgemeinern. So ist es möglich, für jedes Pixel bzw. für jeden Punkt eine eigene Transformation zu berechnen. Wenn zum Beispiel eine Tiefenkarte vorliegt und die Oberflächennormale geschätzt werden kann, kann für jeden Pixel bzw. für jeden Punkt der Punktwolke eine Transformation auf eine Ebene berechnet werden.

[0126] Vorzugsweise werden die von der Kamera aufgenommenen Bilder und/oder die segmentierten Bilder vor dem Entrollen vorverarbeitet. Die Vorverarbeitung umfasst ein Normalisieren des jeweiligen Bildes und optional ein Verbessern der Kanten (zum Beispiel wie oben beschrieben anhand einer Tiefenbildkarte). Das Normalisieren des Bilds kann eine oder mehrere der folgenden Operationen umfassen: Normalisieren der Orientierung des Objekts im Bild durch Rotation; Normalisieren der Position des zu untersuchenden Objekts im Bild (z.B. durch Zuschneiden („Crop“-Operation)), Normalisieren der Größe des Bilds bzw. des Bildzuschnitts (z.B. durch eine Änderung der Größe des Bilds („Resize“-Operation)).

[0127] Fig. 7A bis Fig. 9B zeigen die Ergebnisse einer beispielhaften Bildtransformation anhand von Bildern eines Ankers als Objekt.

[0128] Fig. 7A zeigt ein von der Kamera aufgenommenes Bild des zu inspizierenden Ankers. Fig. 7B zeigt das mit Hilfe des Segmentierungsmodells segmentierte Bild des Ankers. Diese Bilder sind die Eingabe in der Normalisierung bzw. werden einer Normalisierungsoperation unterzogen.

[0129] Fig. 8A bis Fig. 8C zeigen die Ergebnisse der Normalisierungsoperationen (Rotation, Zuschneiden, Skalieren) oder anders ausgedrückt den Ausgang der Normalisierung. Fig. 8A zeigt das normalisierte Kamerabild des Ankers. Fig. 8B zeigt das normalisierte segmentierte Bild des Ankers. Fig. 8C zeigt schematisch die Bestimmung der Achse und der Abstände r_1 und r_2 anhand des in Fig. 8B gezeigten segmentierten Bilds des Ankers.

[0130] Bei der Normalisierung eines Bildes können eine oder mehrere der folgenden Operation durchgeführt werden:

- Schätzung der Orientierung des Ankers und Rotation des Ankers, sodass die Achse des Ankers parallel zur x-Achse des Bildes ist. Dies kann wie oben beschrieben erfolgen;
- Schätzung eines Begrenzungsrahmens des Objekts („BoundingBox“), zum Beispiel mithilfe der Segmentierung, und Zuschneiden („Cropping“) des Bildes, wobei das zugeschnittene Bild dem durch den geschätzten Begrenzungsrahmen definierten Bildausschnitt entspricht. Die Koordinaten des Begrenzungsrahmens bzw. des Bildausschnitts können wie folgt ermittelt werden:
 - $y_{\max}$ = nördlichster Punkt der Segmentierungsmaske für das Objekt,
 - $y_{\min}$ = südlichster Punkt der Segmentierungsmaske für das Objekt,
 - $x_{\max}$ = östlichster Punkt der Segmentierungsmaske für das Objekt,
 - $x_{\min}$ = westlichster Punkt der Segmentierungsmaske für das Objekt;
- Normalisierung der Größe („Resize“) des Bildes, sodass der Bildausschnitt eine vorgegebene Größe bzw. eine vorbestimmte Anzahl von Pixeln (wie z.B. 4.000 Pixeln) enthält.

[0131] Anschließend können wie oben beschrieben die Achse des Objekts und die Abstände r_1 und r_2 aus den normalisierten Bildern geschätzt bzw. ermittelt werden. Beispielsweise können für jede Spalte des zugeschnittenen Bilds die Pixel innerhalb einer Segmentierungsmaske für das Objekt, deren y-Koordinate kleiner (r_1) und größer (r_2) als die Position der Ankerachse ist, gezählt werden. Bei dem in **Fig. 8C** gezeigten Beispiel liegt die Achse in etwa bei $y=100$.

[0132] Anhand der ermittelten Abstände r_1 und r_2 kann das normalisierte Kamerabild und optional das normalisierte segmentierte Bild wie oben beschrieben entrollt werden.

[0133] Wie oben beschrieben, umfasst das Entrollen ein Neuordnen der y-Koordinate im Bild. Die gekrümmte Oberfläche des Ankers wird dabei sinusartig herausgerechnet, um eine Schätzung für die planare Mantelfläche des Ankers, deren Ausschnitt man im Originalbild sieht, zu erhalten.

[0134] Dabei kann zunächst von dem normalisierten Segmentierungsergebnis ausgegangen werden, d.h. die Achse des Objektes ist parallel zur x-Achse im Bild (siehe **Fig. 8C**).

[0135] Dann ist r_1 und r_2 der Abstand in Pixeln zwischen der Achse des Objektes und dem Schnittpunkt der Senkrechten der Achse zur Kante (siehe **Fig. 8C** für r_1). Mithilfe des Radiuses und des Abstandes zum Objekt kann berechnet werden, welcher Anteil der Mantelfläche zu sehen ist.

[0136] Exemplarisch kann angenommen werden, dass die Entfernung so groß ist, sodass der Anteil der Mantelfläche 178° (89° nach oben und nach unten) entspricht. Die y-Koordinate jeder Spalte (d.h. für ein festes x) ist folgendermaßen verzerrt: $\vartheta(x) = \arcsin(y(x) / r_1(x) * \sin(89^\circ))$ für $y > 0$ und $y'(x) = \arcsin(y(x) / r_2(x) * \sin(89^\circ))$ für $y < 0$. (ϑ ist die neue Koordinate des entrollten Bilds).

[0137] **Fig. 9A** und **Fig. 9B** zeigen das Ergebnis des Entrollens der in **Fig. 8A** und **Fig. 8B** gezeigten Bilder, wobei **Fig. 9A** einen Ausschnitt (den mittleren Teil) aus dem entrollten normalisierten Originalbild des Objekts zeigt und **Fig. 9B** einen Ausschnitt (den mittleren Teil) aus dem entrollten normalisierten segmentierten Bild des Objekts zeigt. Diese Bilder werden für die nachfolgende Registrierung verwendet. Der Grund für die Verwendung des mittleren Teils für die Registrierung ist, dass am Rand des jeweiligen Bilds die Verzerrungen aufgrund von kleinen Fehlern bei der Segmentierung nicht mehr vernachlässigt werden können.

Bildregistrierung

[0138] Bildregistrierung ist ein Prozess in der digitalen Bildverarbeitung, welcher dazu dient, mehrere Bilder derselben Szene bzw. desselben Objekts bestmöglich in Übereinstimmung zu bringen. Im Prinzip kann jedes Registrierungsverfahren eingesetzt werden.

[0139] Wie oben beschrieben kann mit Hilfe des Segmentierungsmodells und optional der auf Basis von Kamerabildern ermittelten Lageparameter aus dem hochaufgelösten Bild das Objekt extrahiert und registriert werden. Die aus den Bildern einer zeitlichen Bildabfolge (z.B. einer Videoabfolge) extrahierten und registrierten Objekte werden anschließend fusioniert und in ein Koordinatensystem überführt, das durch den Parameter p und die zur Trajektorie orthogonale Bildkoordinate aufgespannt wird. Eine Verschiebung der Kameraposition um „Op“ führt somit zu einer Translation auf der Ordinatenachse im neuen Koordinatensystem.

[0140] **Fig. 10** zeigt schematisch eine beispielhafte Transformation einer von der Kamera aufgenommenen zeitlichen Abfolge von Bildern eines Objekts (eines Ankers) in einem zwei-dimensionalen Bild, welches den Winkel ϑ und die zur Trajektorie orthogonale Bildkoordinate als Ortskoordinaten aufweist. Die zeitliche Abfolge von Bildern ist von einer beweg-

lichen Kamera aufgenommen, welche sich auf der in **Fig. 2B** gezeigten Trajektorie mit dem Winkel ϑ als Parameter bewegt. **Fig. 10A** zeigt schematisch ein der aufgenommenen Bilder und den Abstand $r(x)$ (für $y < 0$). **Fig. 10B** zeigt das Ergebnis der Transformation.

[0141] Nachfolgend werden zwei beispielhafte Registrierungsverfahren beschrieben.

Erstes beispielhaftes Registrierungsverfahren

[0142] Für die Registrierung werden transformierte (z.B. normalisierte und entrollte) Kamerabilder verwendet (siehe z.B. **Fig. 9A**). Zunächst werden in jedem transformierten Bild Merkmale an bestimmten Punkten („Points-of-interest“) berechnet. Die Punkte und die Merkmale können mittels bekannter Verfahren, z.B. durch ORB (siehe z.B. „ORB: an efficient alternative to SIFT or SURF“, Proceedings / IEEE International Conference on Computer Vision, IEEE International Conference on Computer Vision, doi:10.1109/ICCV.2011.6126544) oder SIFT (siehe z.B. D. G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2):91-110, 2004), ermittelt werden. Anschließend werden die Punkte von jeweils zwei aufeinanderfolgenden Bildern gematcht („feature matching“). Dies kann durch das Finden von Punktepaaren mit der niedrigsten euklidischen Distanz erfolgen. Anhand der aufgefundenen Punktepaare kann eine euklidische Transformationsmatrix (zum Beispiel eine affine Transformation ohne Scherung) zwischen den beiden aufeinanderfolgenden Bildern berechnet werden. Die beiden Bilder werden anschließend zusammengesetzt, d.h. fusioniert, indem ein Bild mithilfe der Transformationsmatrix in Deckung zum anderen Bild gebracht wird. Im Bereich der Überschneidung wird der Teil des jeweiligen Bildes genommen, der näher im Zentrum der jeweiligen Bild ist, d.h. dort, wo $|p|$ (z.B. $|p|$) am niedrigsten ist, da dort Verzerrungsfehler der Transformation am geringsten sind.

Zweites beispielhaftes Registrierungsverfahren

[0143] Bei dem zweiten beispielhaften Registrierungsverfahren kann bei der Registrierung das Segmentierungsergebnis durch eine Schätzung der Bodenebene und/oder durch die Verwendung eines semantischen Objektmodells (wie z.B. des in den Publikationen [8] und [9] beschriebenen semantischen Objektmodells) verbessert werden. Wie oben beschrieben, wird mit Hilfe des Segmentierungsergebnisses die Größe und Orientierung des Objektes normalisiert, damit das Objekt in jedem Bild die gleiche Größe und Orientierung hat. Anschließend findet eine Transformation (z.B. Entrollen) statt. Mit dieser Vorverarbeitung reduziert sich die Registrierung auf eine Suche nach der besten Translation. Hierzu kann

Template-Matching verwendet werden. Für jede (sinnvolle) Verschiebung in x- und y-Richtung des zweiten Bildes wird eine Korrelation mit dem nicht verschobenen ersten Bild berechnet. Die beste Translation ergibt sich aus der Position des Maximums der Korrelationsfunktion. Anschließend erfolgt eine Fusion der Bilder, wie in Zusammenhang mit der ersten Option beschrieben.

[0144] **Fig. 11** zeigt ein beispielhaftes Verfahren zum Registrieren von Bildern aus der zeitlichen Abfolge von Bildern mittels Template Matching, wobei ein Modell M1 verwendet wird. Das Modell entspricht den bereits zusammengesetzten transformierten Kamerabildern. Falls das aktuelle Kamerabild das erste ist, wird dieses als initiales Modell für die folgenden Kamerabilder verwendet. In einem ersten Schritt S1 wird eine neue Kamerabild als eine neue Eingabe erfasst. Das Kamerabild wird in einem zweiten Schritt S2 transformiert. Das Kamerabild kann zum Beispiel wie oben beschrieben normalisiert und entrollt werden. In einem dritten Schritt S3 wird ein Muster- bzw. Schablonenvergleich („Template-Matching“) zwischen dem transformierten Bild und dem Modell M1 durchgeführt. In einem vierten Schritt S4 wird entschieden, ob das Ergebnis des Muster- bzw. Schablonenvergleichs über einem vorgegebenen Schwellenwert „t“ liegt. Wenn das Ergebnis des Muster- bzw. Schablonenvergleichs über einen vorgegebenen Schwellenwert „t“ liegt, wird der „Ja“ Pfad genommen und das Modell M1 mit der aktuellen Eingabe fusioniert (Schritt S5), d.h. das Modell wird um einen Teil der aktuellen Eingabe erweitert (M2). Das so aktualisierte Modell M2 wird als Eingabemodell für das nächste Eingangsbild verwendet. Wenn das Ergebnis des Muster- bzw. Schablonenvergleichs unter dem vorgegebenen Schwellenwert „t“ liegt, wird der Pfad „Nein“ genommen und auf eine neue Eingabe gewartet (Schritt S6).

[0145] Während beim oben beschriebenen ersten Registrierungsverfahren aufeinander folgende entrollte Bilder mit ORB-Merkmalen abgeglichen werden, um übereinstimmende Regionen zu finden, wird beim zweiten Verfahren ein Schablonenabgleich verwendet, um das Modell der Mantelfläche zu vergrößern.

[0146] Ein Nachteil bei ORB-Merkmalen bei manchen Objekten (wie zum Beispiel Ankern) kann sein, dass die Oberflächentextur annähernd periodisch ist und eine relativ große Anzahl von falschen Punkten gematcht werden können. Ein Schablonenvergleich (Template Matching) funktioniert besser aufgrund kleiner Fehler auf der Oberfläche und weil einzelne Elemente (wie z.B. die Kupferdrähte eines Ankers) einen stärkeren Einfluss haben.

Defekterkennung

[0147] Zur Defekterkennung können unterschiedliche an sich bekannte Verfahren verwendet werden. Beispielsweise können folgende Verfahren verwendet werden:

1. Novelty Detection mit CNNs (CNN: Convolutional Neural Network bzw. faltendes neuronales Netzwerk) oder eines anderen geeigneten neuronalen Netzes: Hierbei wird ein Autoencoder mit nicht annotierten Bildern trainiert. Da Defekte in der Regel äußerst selten auftreten können und ihr Erscheinungsbild variieren kann, kann der Autoencoder die Defekte in seinem Merkmalsraum nicht darstellen. In einem Beispiel kann der Autoencoder daher nur die defektfreien Regionen konstruieren. Das Differenzbild zwischen dem Originalbild und dem rekonstruierten Bild gibt Aufschluss darüber, wo und ob sich Defekte im Bild befinden. Für eine binäre Klassifikation kann ein Schwellwertverfahren hinsichtlich Größe und/oder Amplitude der Defekte festgelegt werden.

2. Spektralanalyse: Defekte in periodischen Texturen (bei einem Anker wäre dies beim Kommutator gegeben) können im Spektrogramm des Bildes erkannt werden. Hierzu wird mithilfe des Segmentierungsergebnisses die Textur extrahiert (z.B. durch Fensterung) und Fouriertransformiert. Die Textur ist dann durch Spitzen („peaks“) im Spektrogramm identifizierbar. Spektralanteile mit einer kleineren Amplitude sind Störungen. Durch eine geeignete Vorverarbeitung der Bilder (umfassend zum Beispiel Entfernen von Bildrauschen und/oder Schwankungen bei der Beleuchtung) und ein geeignetes Schwellwertverfahren hinsichtlich Größe und/oder Amplitude der Defekte können Defekte erkannt werden. Auch Defekte, die sich durch eine Vorzugsrichtung im Bild auszeichnen (z.B. Schleifriefen), lassen sich mit dieser Vorgehensweise erkennen.

3. Deflektometrie bzw. Streifenlichtprojektion für spiegelnde Oberflächen: Durch Projektion eines bestimmten Musters auf das Werkstück können, falls die Reflexion nicht mit der erwarteten Reflexion übereinstimmt, sowohl Fehler in der Geometrie des Werkstücks (Beulen, Dellen) als auch Verschmutzungen/ Defekte durch schlechte Reflexionseigenschaften abgeleitet werden.

[0148] Die obigen Defekterkennungsverfahren sind lediglich beispielhaft. Es ist möglich, andere geeignete Defekterkennungsverfahren einzusetzen.

Bezugszeichenliste

10

Objekt

12	Kamera
14	Trajektorie
16	Roboterarm
18	Segmentierungsmaske
20	Achse
r1, r2	Abstände von der Achse zu einer Kante des Objekts im Bild
M1-M2	Eingabemodell und aktualisiertes Modell
T	Template
S1-S6	Verfahrensschritte

Patentansprüche

1. Vorrichtung zur optischen Oberflächeninspektion von Objekten umfassend:

eine bewegbare Kamera (12), welche auf einer mit einem Parameter p parametrisierbare Trajektorie (14) um ein zu inspizierendes Objekt (10) bewegbar ist, wobei die Kamera (12) eingerichtet ist, während ihrer Bewegung auf der Trajektorie eine zeitliche Abfolge von Bildern des zu inspizierenden Objekts (10) aus unterschiedlichen Positionen auf der Trajektorie (14) aufzunehmen;

eine Rechenvorrichtung umfassend:

eine Modellbereitstellungseinheit, welche eingerichtet ist, ein statistisches Segmentierungsmodell bereitzustellen, wobei das Segmentierungsmodell anhand einer Mehrzahl von Trainingsbildern des Objekts (10) in unterschiedlichen Zuständen trainiert worden ist, wobei das Segmentierungsmodell als Eingang ein Bild nimmt und als Ergebnis eine Klassifikation der Pixel im Eingangsbild in zumindest einer Klasse liefert, wobei die zumindest eine Klasse einer zu inspizierenden (Teil-)Oberfläche des Objekts (10) entspricht;

eine Bildsegmentierungseinheit, die eingerichtet ist, ein von der Kamera (12) aufgenommenes Bild mit Hilfe des Segmentierungsmodells zu segmentieren; eine Trajektorienbestimmungseinheit, die eingerichtet ist, die Trajektorie (14) anhand zumindest eines von der Kamera (12) aufgenommenen und von der Bildsegmentierungseinheit segmentierten Bilds des Objekts (10) zu bestimmen;

eine Bildtransformationseinheit, die eingerichtet ist, aus jedem Bild der zeitlichen Abfolge von Bildern, welche von der Kamera (12) bei ihrer Bewegung auf der Trajektorie (14) aufgenommen worden sind, zumindest eine zu inspizierende (Teil-)Oberfläche des Objekts (10) anhand des Segmentierungsergebnis für das Bild zu extrahieren und die extrahierten (Teil-)Oberflächen in einer zwei-dimensionalen Darstellung zu fusionieren, wobei die zwei-dimensionale Darstellung den Parameter p und eine zur Trajektorie (14) orthogonale Bildkoordinate als Ortskoordinaten aufweist;

eine Defekterkennungseinheit, die eingerichtet ist, eine Defekterkennung auf die fusionierte zwei-dimensionale Darstellung durchzuführen.

2. Vorrichtung nach Anspruch 1, wobei die Trajektorien-bestimmungseinheit eingerichtet ist, eine Hauptkomponentenanalyse des segmentierten Bilds durchzuführen, eine Achse des Objekts (10) anhand des Ergebnisses der Hauptkomponentenanalyse zu bestimmen, und die Trajektorie (14) derart zu bestimmen, dass die auf der Trajektorie (14) liegenden Punkte einen vorbestimmten oder vorbestimmbaren Abstand zur Achse aufweisen.

3. Vorrichtung nach Anspruch 1 oder 2, wobei die Rechenvorrichtung ferner eingerichtet ist, eine Tiefenkarte des Objekts (10) bereitzustellen; aus der Tiefenkarte die Tiefenkanten aller Objekte, die sich von einer Bodenebene erheben, zu bestimmen, und die Kanten des Objekts (10) mit dem größten Schnitt mit dem Objekt (10) aus der Segmentierung durch die ermittelten Tiefenkanten zu verbessern.

4. Vorrichtung nach einem der vorangegangenen Ansprüche, wobei die Rechenvorrichtung ferner eine Kameraorientierungseinheit umfasst, welche eingerichtet ist, eine Orientierung der Kamera (12) zum Objekt (10) anhand zumindest eines von der Kamera (12) aufgenommenen und von der Bildsegmentierungseinheit segmentierten Bilds des Objekts (10) zu bestimmen.

5. Vorrichtung nach einem der vorangegangenen Ansprüche, wobei die Bildtransformationseinheit eingerichtet ist, vor dem Extrahieren eine Transformation jedes der von der Kamera (12) aufgenommenen Bilder und der entsprechenden segmentierten Bilder durchzuführen, wobei die Transformation umfasst:
Entrollen des jeweiligen Bilds umfassend Projizieren der zumindest einen zu inspizierenden (Teil-)Oberfläche des Objekts (10) auf eine planare Ebene; und/oder
Kompensieren der perspektivischen Verzerrung des jeweiligen Bilds; und/oder
Normalisieren des jeweiligen Bilds.

6. Vorrichtung nach Anspruch 5, wobei das Entrollen des von der Kamera (12) aufgenommenen Bilds und des entsprechenden segmentierten Bilds umfasst:
Bestimmen aus dem segmentierten Bild einer Achse des Objekts (10) mittels Hauptkomponentenanalyse;
Bestimmen zumindest eines Abstands von der Achse zu einer Kante des Objekts (10) im segmentierten Bild;
Bestimmen einer Koordinatentransformation anhand

des zumindest einen Abstands;

Transformieren des von der Kamera (12) aufgenommenen Bilds und des entsprechenden segmentierten Bilds anhand der Koordinatentransformation,

7. Vorrichtung nach einem der vorangegangenen Ansprüche, wobei das Fusionieren der extrahierten Bilder in der zwei-dimensionalen Darstellung umfasst:

für jedes Paar von zwei aufeinanderfolgenden Bildern, die von der Kamera (12) aufgenommen und optional transformiert worden sind:

Bestimmen von einer Mehrzahl von Merkmalen an bestimmten Punkten in jedem der Bilder;

Auffinden von Punktpaaren mit übereinstimmenden Merkmalen in den zwei aufeinanderfolgenden Bildern;

Bestimmen einer Transformationsmatrix zwischen den aufeinander folgenden Bildern;

Fusionieren der aufeinanderfolgenden Bilder, indem eines der Bilder mit Hilfe der Transformationsmatrix in Deckung zum anderen Bild gebracht wird, wobei im Bereich der Überschneidung der Teil des jeweiligen Bildes genommen wird, der näher am Zentrum des jeweiligen Bilds ist.

8. Vorrichtung nach einem der Ansprüche 1 bis 6, wobei das Fusionieren der extrahierten Bilder in der zwei-dimensionalen Darstellung umfasst:

für jedes Paar von zwei aufeinanderfolgenden Bildern, die von der Kamera (12) aufgenommen und optional transformiert worden sind:

Auffinden einer Translation eines der Bilder gegenüber dem anderen Bild, die eine Korrelationsfunktion zwischen den beiden Bildern maximiert; und

Fusionieren der aufeinanderfolgenden Bilder, indem eines der Bilder mit Hilfe der Translation in Deckung zum anderen Bild gebracht wird, wobei im Bereich der Überschneidung der Teil des jeweiligen Bildes genommen wird, der näher am Zentrum des jeweiligen Bilds ist.

9. Computerimplementiertes Verfahren zur optischen Oberflächeninspektion von Objekten umfassend:

Bereitstellen eines statistischen Segmentierungsmodells, wobei das Segmentierungsmodell anhand einer Mehrzahl von Trainingsbildern eines Objekts (10) in unterschiedlichen Zuständen trainiert worden ist, wobei das Segmentierungsmodell als Eingang ein Bild nimmt und als Ergebnis eine Klassifikation der Pixel im Eingangsbild in zumindest einer Klasse liefert, wobei die zumindest eine Klasse einer zu inspizierenden (Teil-)Oberfläche des Objekts (10) entspricht.

Bestimmen einer mit einem Parameter p parametrisierbaren Trajektorie (14), um ein zu inspizierendes Objekt anhand zumindest eines von einer Kamera (12) aufgenommenen und mit Hilfe des Segmentierungsmodells segmentierten Bilds des Objekts (10);

Aufnahme einer zeitlichen Abfolge von Bildern des zu inspizierenden Objekts (10) von der Kamera (12), wobei sich die Kamera auf der Trajektorie (14) um das Objekt (10) bewegt und während ihrer Bewegung Bilder des Objekts (10) aus unterschiedlichen Positionen auf der Trajektorie (14) aufnimmt; Segmentieren der Bilder der zeitlichen Abfolge von Bildern (10) mit Hilfe des Segmentierungsmodells; Extrahieren zumindest einer zu inspizierenden (Teil-)Oberfläche des Objekts (10) aus jedem der von der Kamera (12) aufgenommenen Bilder mit Hilfe des Segmentierungsergebnis für das jeweilige Bild; Fusionieren der extrahierten (Teil-)Oberflächen in einer zwei-dimensionalen Darstellung, wobei die zwei-dimensionale Darstellung den Parameter p und eine zur Trajektorie (14) orthogonale Bildkoordinate als Ortskoordinaten aufweist; und Durchführen einer Defekterkennung auf die fusionierte zwei-dimensionale Darstellung.

10. Verfahren nach Anspruch 9, wobei das Bestimmen der Trajektorie (14) umfasst: Bestimmen einer Achse des Objekts (10) anhand einer Hauptkomponentenanalyse des segmentierten Bilds, und Bestimmen der Trajektorie (14) derart, dass die auf der Trajektorie (14) liegenden Punkte einen vorbestimmten oder vorbestimmbaren Abstand zur Achse aufweisen.

11. Verfahren nach Anspruch 9 oder 10, ferner umfassend: Bereitstellen einer Tiefenkarte des Objekts (10); Bestimmen von Tiefenkanten aller Objekte, die sich von einer Bodenebene erheben anhand der Tiefenkarte, und Verbessern der Kanten des Objekts (10) mit dem größten Schnitt mit dem Objekt (10) aus der Segmentierung mit Hilfe der Tiefenkanten.

12. Verfahren nach einem der Ansprüche 9 bis 11, ferner umfassend Bestimmen einer Orientierung der Kamera (12) zum Objekt (10) anhand zumindest eines von der Kamera (12) aufgenommenen und mit Hilfe des Segmentierungsmodells segmentierten Bilds des Objekts (10).

13. Verfahren nach einem der Ansprüche 9 bis 12, ferner umfassend Transformieren jedes der von der Kamera (12) aufgenommenen Bilder und der entsprechenden segmentierten Bilder vor dem Extrahieren, wobei das Transformieren umfasst: Entrollen des jeweiligen Bilds umfassend Projizieren der zumindest einen zu inspizierenden (Teil-)Oberfläche des Objekts (10) auf eine planare Ebene; und/oder Kompensieren der perspektivischen Verzerrung des jeweiligen Bilds; und/oder Normalisieren des jeweiligen Bilds.

14. Verfahren nach Anspruch 13, wobei das Entrollen des von der Kamera (12) aufgenommenen Bilds und des entsprechenden segmentierten Bilds durchzuführen umfasst:

Bestimmen einer Achse des Objekts (10) mit Hilfe einer Hauptkomponentenanalyse des segmentierten Bilds; Bestimmen zumindest eines Abstands von der Achse zu einer Kante des Objekts (10) im segmentierten Bild; Bestimmen einer Koordinatentransformation anhand des zumindest einen Abstands; und Transformieren des von der Kamera (12) aufgenommenen Bilds und des entsprechenden segmentierten Bilds anhand der Koordinatentransformation.

15. Verfahren nach einem der Ansprüche 9 bis 14, wobei das Fusionieren der extrahierten Bilder in der zwei-dimensionalen Darstellung umfasst: für jedes Paar von zwei aufeinanderfolgenden Bildern, die von der Kamera (12) aufgenommen und optional transformiert worden sind: Bestimmen von einer Mehrzahl von Merkmalen an bestimmten Punkten in jedem der Bilder; Auffinden von Punktpaaren mit übereinstimmenden Merkmalen in den zwei aufeinanderfolgenden Bildern; Bestimmen einer Transformationsmatrix zwischen den aufeinander folgenden Bildern; Fusionieren der aufeinanderfolgenden Bilder, indem eines der Bilder mit Hilfe der Transformationsmatrix in Deckung zum anderen Bild gebracht wird, wobei im Bereich der Überschneidung der Teil des jeweiligen Bildes genommen wird, der näher am Zentrum des jeweiligen Bilds ist.

16. Verfahren nach einem der Ansprüche 9 bis 14, wobei das Fusionieren der extrahierten Bilder in der zwei-dimensionalen Darstellung umfasst: für jedes Paar von zwei aufeinanderfolgenden Bildern, die von der Kamera (12) aufgenommen und optional transformiert worden sind: Auffinden einer Translation eines der Bilder gegenüber dem anderen Bild, die eine Korrelationsfunktion zwischen den beiden Bildern maximiert; und Fusionieren der aufeinanderfolgenden Bilder, indem eines der Bilder mit Hilfe der Translation in Deckung zum anderen Bild gebracht wird, wobei im Bereich der Überschneidung der Teil des jeweiligen Bildes genommen wird, der näher am Zentrum des jeweiligen Bilds ist.

17. Computerprogrammprodukt, welches eingerichtet ist, wenn geladen und ausgeführt auf einem Computer, ein Verfahren zur optischen Oberflächeninspektion von Objekten (10) nach einem der Ansprüche 9 bis 16 durchzuführen.

Es folgen 11 Seiten Zeichnungen

Anhängende Zeichnungen

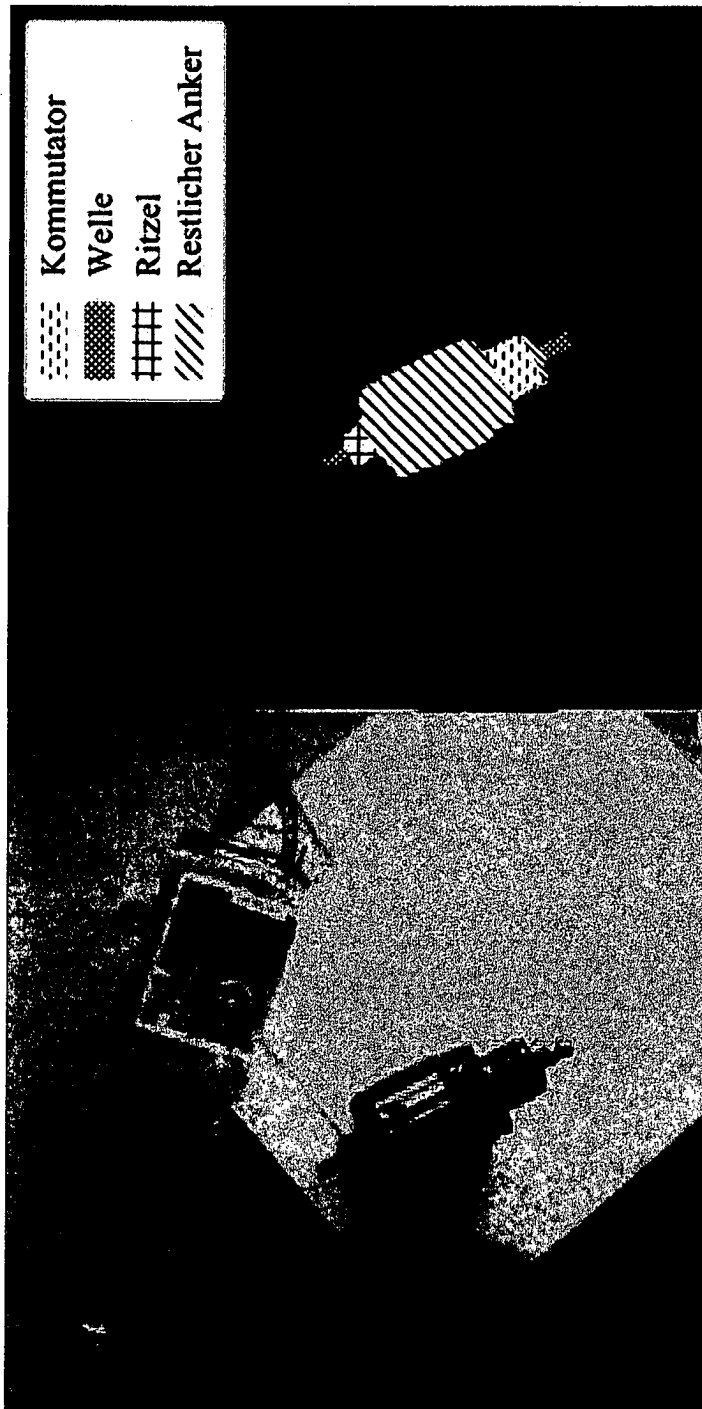


Fig. 1B

Fig. 1A

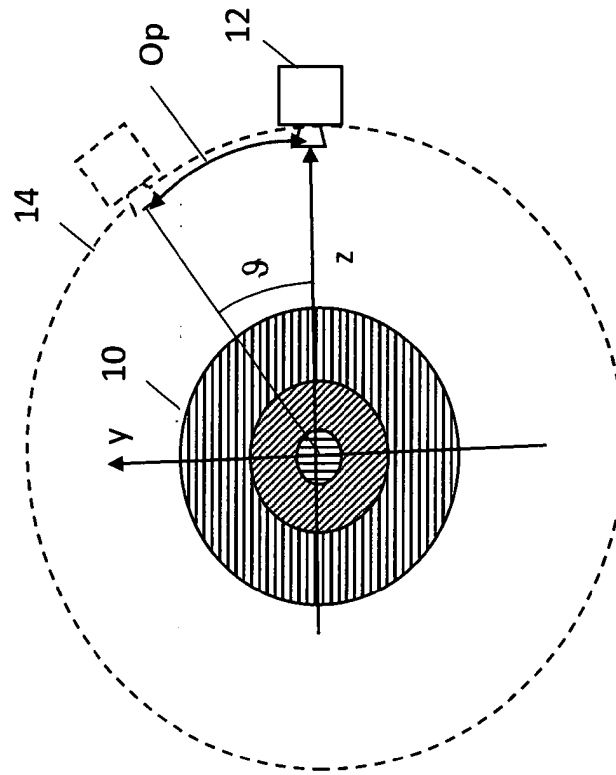


Fig. 2B

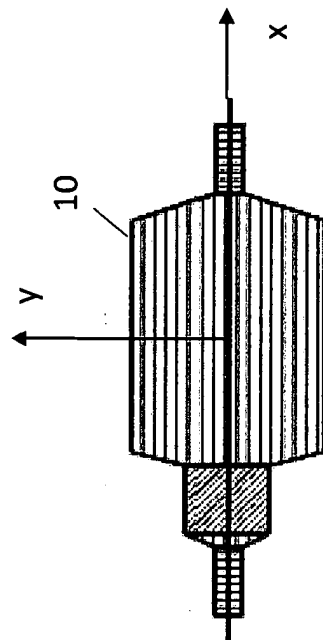


Fig. 2A

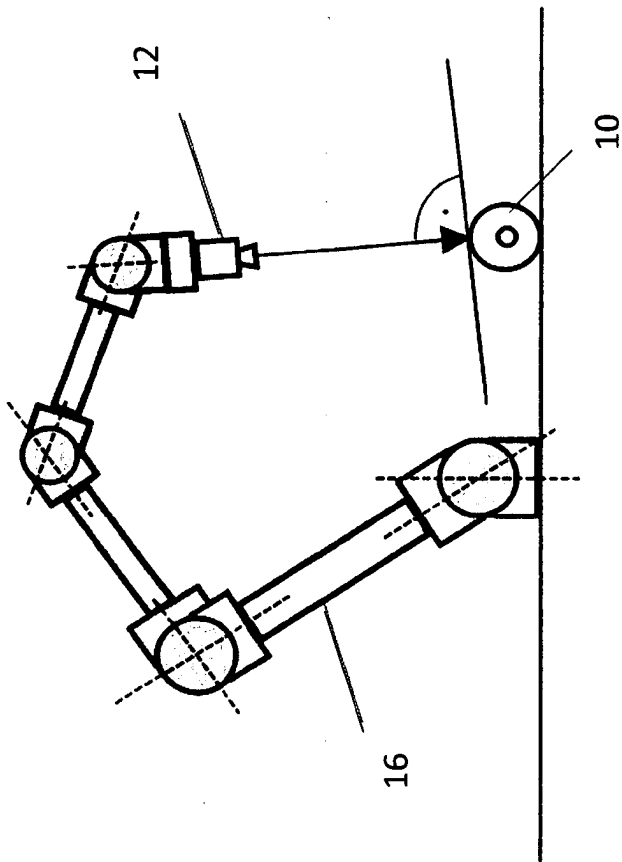


Fig. 3A

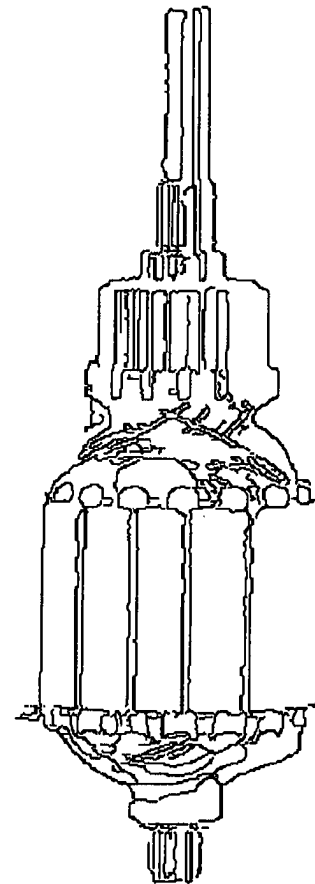


Fig. 3B



Fig. 4B

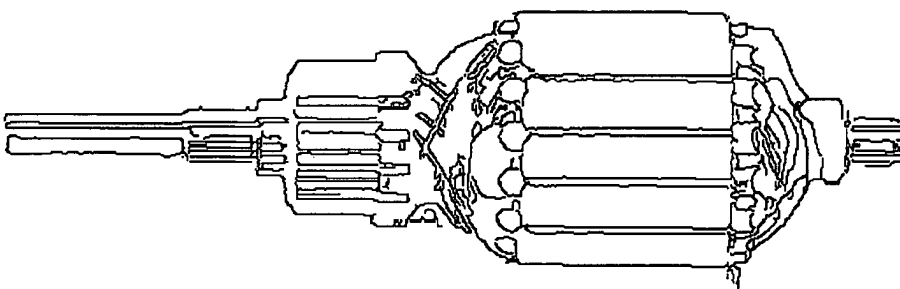


Fig. 4A

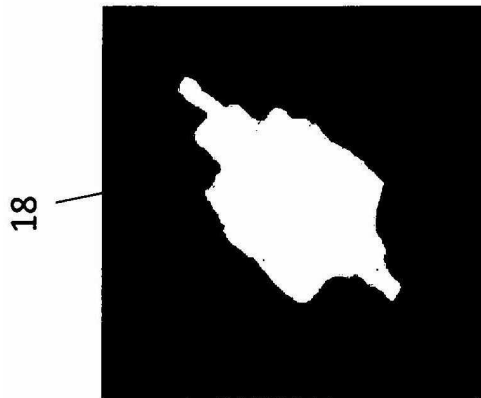


Fig. 5A

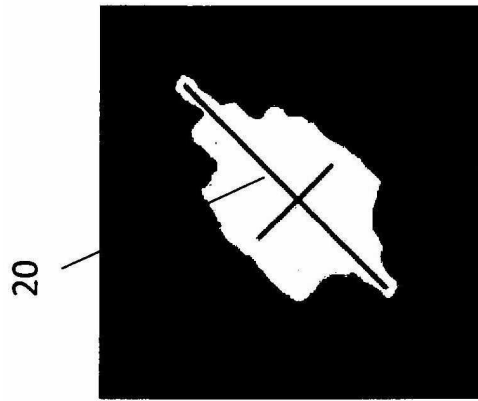


Fig. 5B

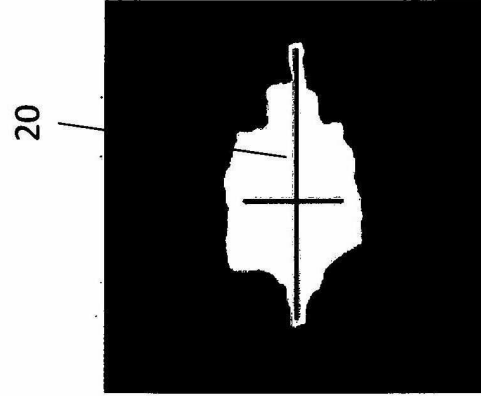


Fig. 5C

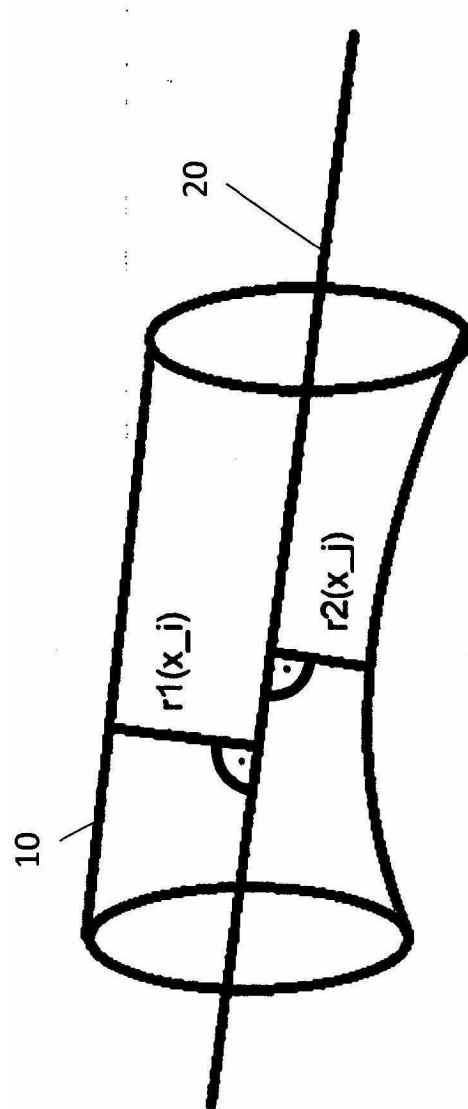


Fig. 6

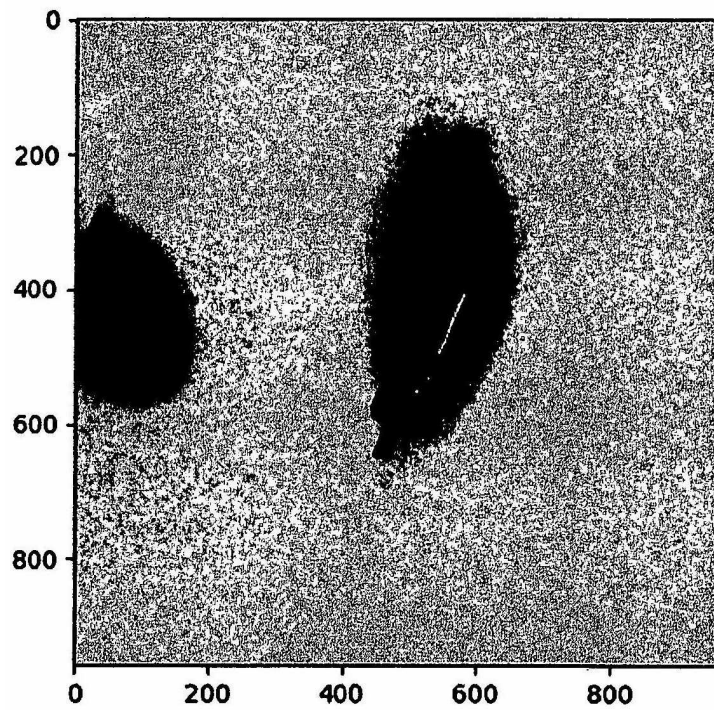


Fig. 7A

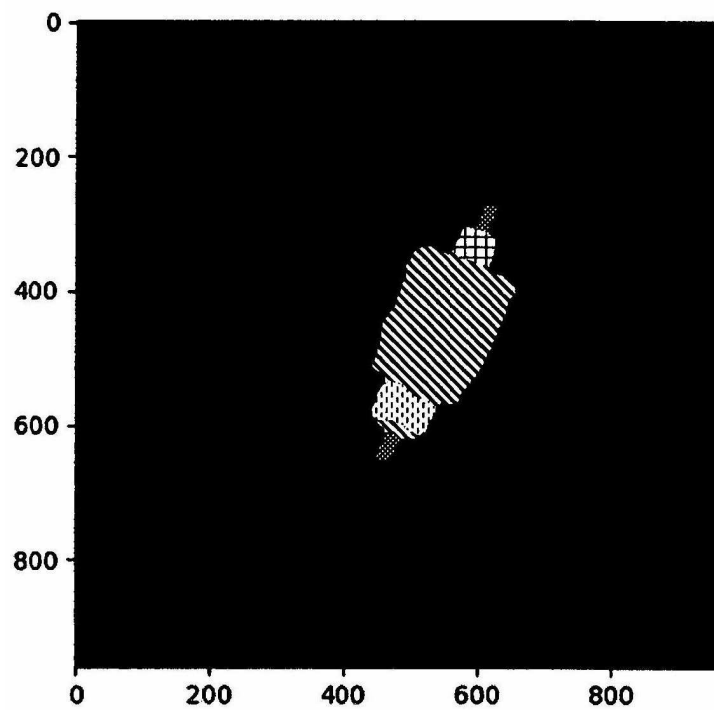


Fig. 7B

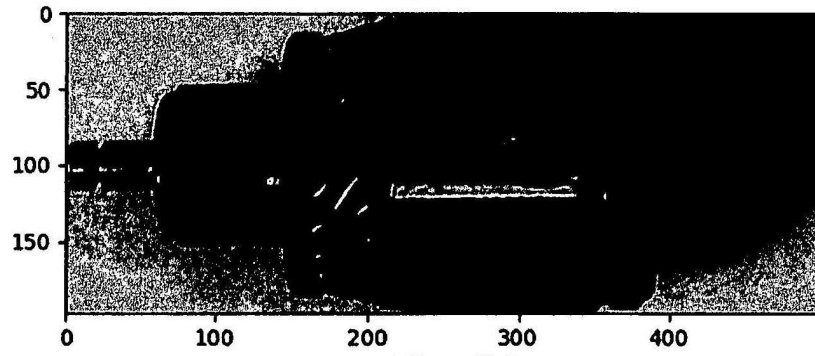


Fig. 8A

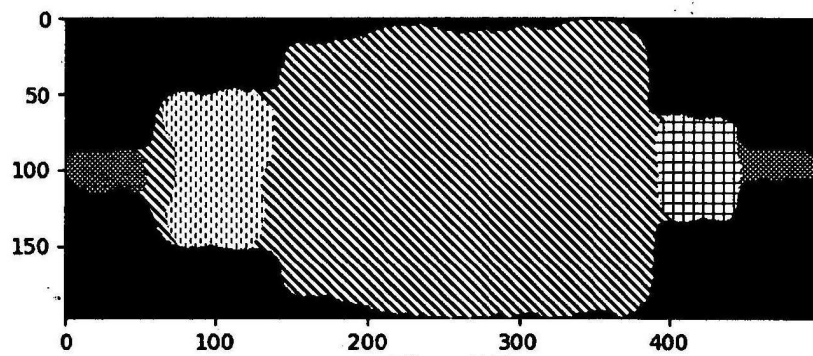


Fig. 8B

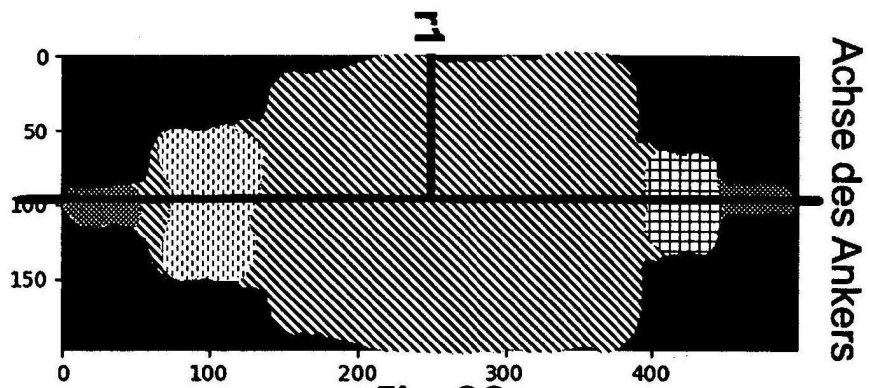


Fig. 8C

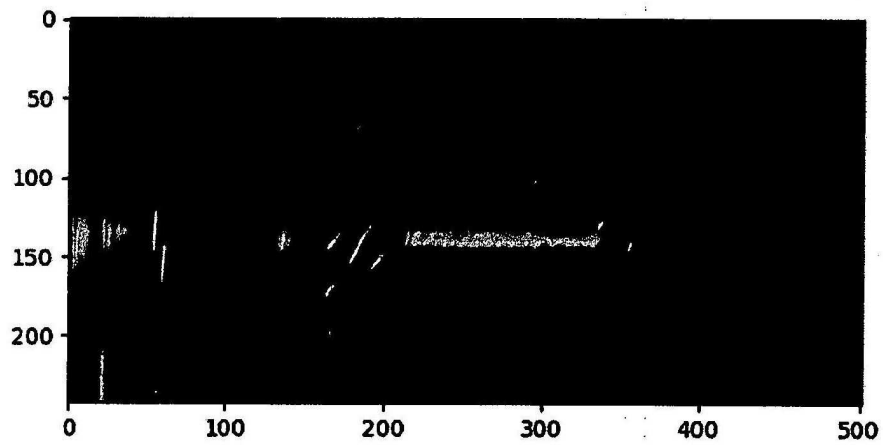


Fig. 9A

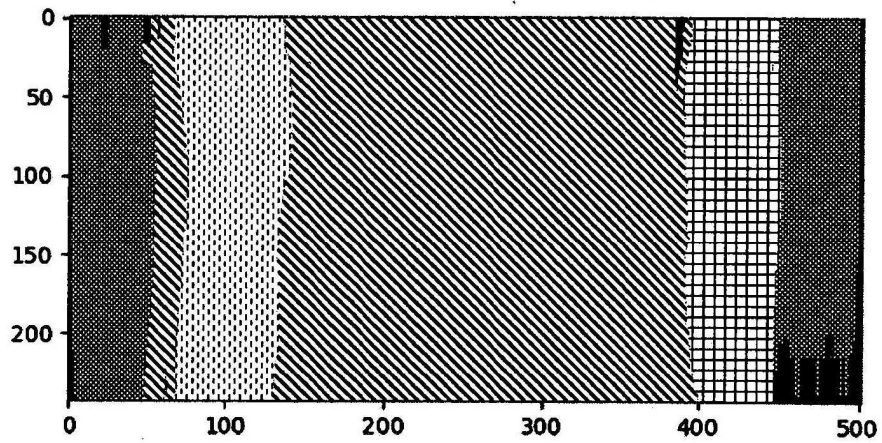


Fig. 9B

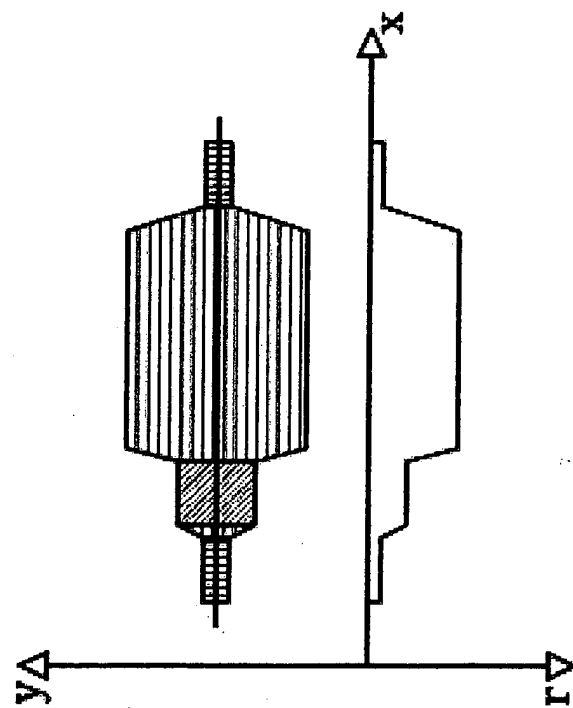


Fig. 10A

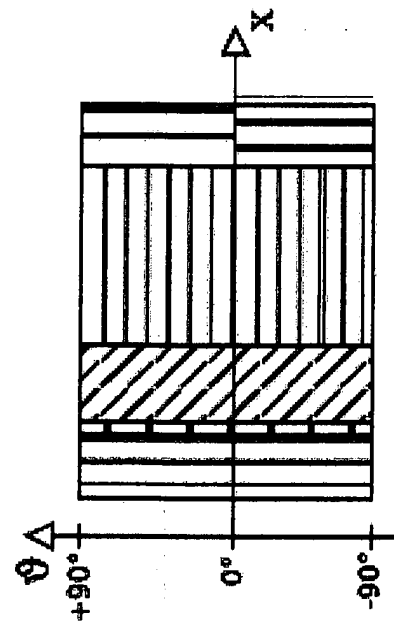


Fig. 10B

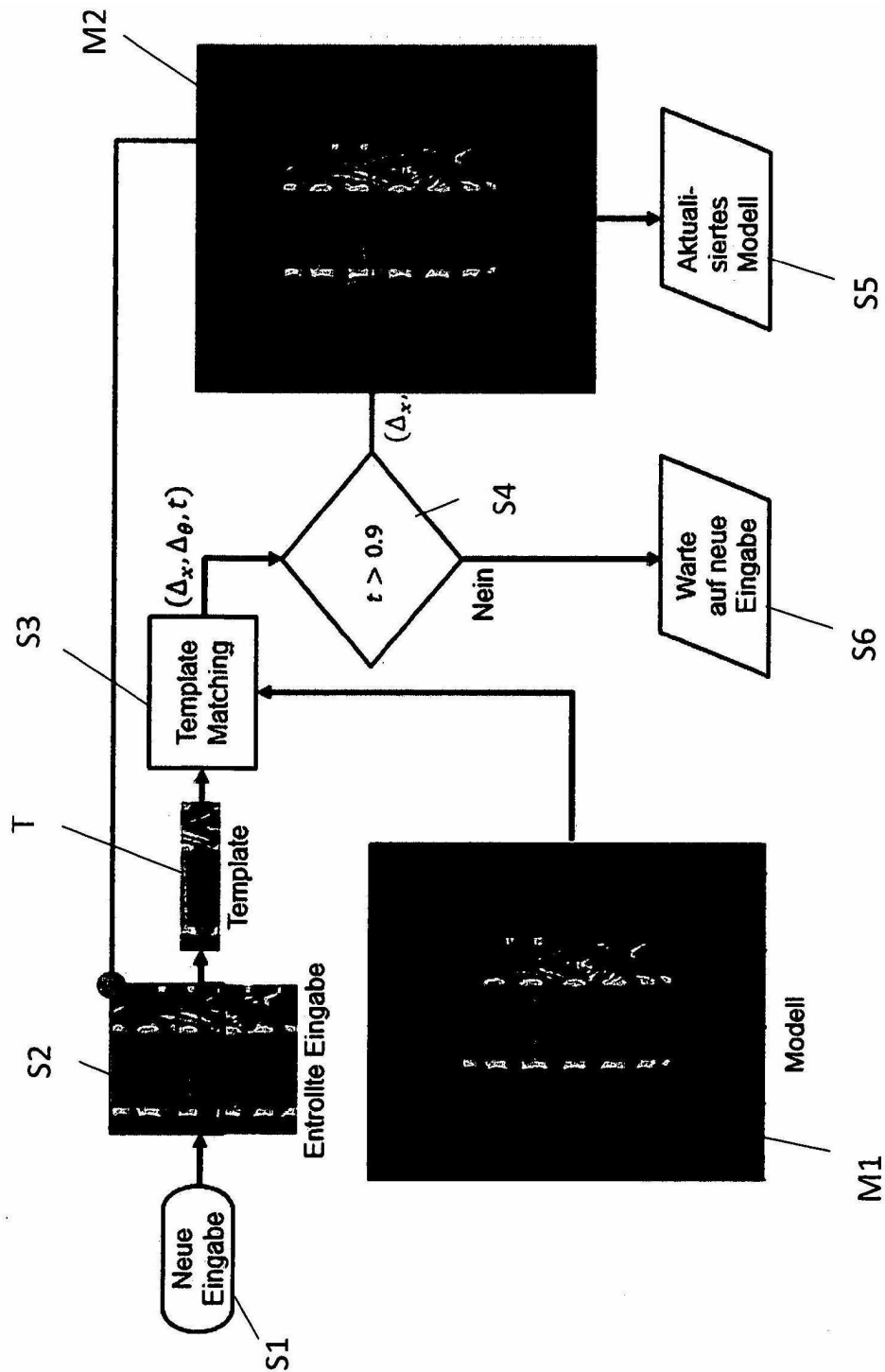


Fig. 11