



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 10 2009 031 804 A1** 2010.03.11

(12)

Offenlegungsschrift

(21) Aktenzeichen: **10 2009 031 804.6**

(22) Anmeldetag: **06.07.2009**

(43) Offenlegungstag: **11.03.2010**

(51) Int Cl.⁸: **G06K 9/46** (2006.01)
G06K 9/62 (2006.01)

Mit Einverständnis des Anmelders offengelegte Anmeldung gemäß § 31 Abs. 2 Ziffer 1 PatG

(71) Anmelder:
Daimler AG, 70327 Stuttgart, DE

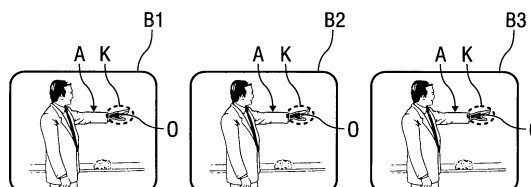
(72) Erfinder:
Hahn, Markus, Dipl.-Inform., 99718 Westgreußen, DE; Krüger, Lars, Dr., 89073 Ulm, DE; Quronfuleh, Fuad, Dipl.-Ing. (FH), 89081 Ulm, DE; Wöhler, Christian, Dr.rer.nat., 72535 Heroldstatt, DE

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

(54) Bezeichnung: **Verfahren zur Objekterkennung und Objektverfolgung**

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zur Objekterkennung und Objektverfolgung unter Verwendung von mittels einer Kameraanordnung erfassten Bildern (B1, B2, B3), wobei die Kameraanordnung eine Mehrzahl von Kameras umfasst, wobei mittels einer Verarbeitungseinheit aus den Bildern (B1, B2, B3) der Kameras ein dreidimensionales Bewegungsfeld erstellt wird, wobei zumindest ein Objekt (O) erkannt und über einen Zeitverlauf verfolgt wird.

Erfindungsgemäß wird das zumindest eine Objekt (O) mittels zumindest eines Klassifizierungsverfahrens erkannt und Disparitätsbildpunkte des zumindest einen Objekts (O) werden zu zumindest einem Cluster zusammengefasst, wobei der zumindest eine Cluster des zumindest einen erkannten Objekts (O) durch zumindest einen vorgegebenen einfachen geometrischen Körper (K) approximiert wird.



Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zur Objekterkennung und Objektverfolgung nach den Merkmalen des Oberbegriffs des Anspruchs 1.

[0002] Aus dem Stand der Technik ist, wie in der US 2003/0081815 A1 beschrieben, eine Entfernungsbestimmungseinheit bekannt, welche eine Bildverarbeitung verwendet. Ein Bildgrößenänderungsbereich erzeugt ein größenverändertes Bild, indem die Größe eines von zwei Originalbildern, welche mittels eines Kameraaars erfasst wurden, verändert wird. Wenn eine Kante eines Objekts mehrere schräge Bestandteile aufweist, ist die Kante schwierig als vertikale Kante zu ermitteln. Wird eine Größe des Bildes horizontal reduziert, nähert sich die schräge Kante einer vertikalen Kante an. Dadurch sind in einem Charakteristikextrahierungsbereich charakteristische Endpunkte zuverlässig extrahierbar, so dass ein Objekt erkannt wird und eine Entfernung zu diesem Objekt zuverlässig bestimmt werden kann.

[0003] Des Weiteren ist eine Objekterkennung und Objektverfolgung bekannt aus D. Comaniciu, V. Ramesh, P. Meer: Real-Time Tracking of Non-Rigid Objects using Mean Shift, IEEE Conf. Computer Vision and Pattern Recognition (CVPR'00), Hilton Head Island, South Carolina, Vol. 2, pp. 142–149, 2000, sowie aus D. Comaniciu, V. Ramesh, P. Meer: Kernel-Based Object Tracking, IEEE Trans. Pattern Analysis Machine Intell., Vol. 25, No. 5, pp. 564–575, 2003 und aus G. R. Gradscki, Real time face and object tracking as a component of a perceptual user interface. In WACV'98: Proceedings of the 4th IEEE Workshop on Applications of Computer Vision, pp. 214–219, Washington, DC, USA. IEEE Computer Society. Bei dieser Objekterkennung und Objektverfolgung mittels des so genannten Mean-Shift-Tracking werden Bildmerkmale wie Farben, Helligkeit, Intensitäten oder Kanten benutzt, um Ähnlichkeiten zwischen einem vorher definierten Referenzobjekt und einem Bild zu finden. Die Merkmale werden dabei durch Histogramme repräsentiert, was diese Verfahren sehr effektiv macht und zugleich zu einer robusten und zeitlich stabilen Verfolgung führt. Dabei wird das zu verfolgende Objekt durch seine statistische Farbverteilung charakterisiert, welche durch ein Histogramm beschrieben wird. Die Referenzobjekte bzw. deren Farbmodell werden manuell vorgegeben. Mit Hilfe der Mean-Shift-Optimierung wird die Ähnlichkeit zwischen Zielobjekt und Eingabebild maximiert und somit das Referenzobjekt über die Zeit verfolgt. Die beschriebenen Mean-Shift-Trackingverfahren arbeiten auf Basis von zweidimensionalen Bildkoordinaten, so dass eine Entfernungsänderung des verfolgten Objekts sich in einer Größenänderung (Skalierung) der geometrischen Form widerspiegelt.

[0004] Der Erfindung liegt die Aufgabe zugrunde,

ein verbessertes Verfahren zur Objekterkennung und Objektverfolgung anzugeben.

[0005] Die Aufgabe wird erfindungsgemäß durch ein Verfahren zur Objekterkennung und Objektverfolgung mit den Merkmalen des Anspruchs 1 gelöst.

[0006] Bevorzugte Ausgestaltungen und Weiterbildungen der Erfindung sind in den abhängigen Ansprüchen angegeben.

[0007] In einem Verfahren zur Objekterkennung und Objektverfolgung unter Verwendung von mittels einer Kameraanordnung erfasster Bilder, wobei die Kameraanordnung eine Mehrzahl von Kameras umfasst, wobei mittels einer Verarbeitungseinheit aus den Bildern der Kamera ein dreidimensionales Bewegungsfeld erstellt wird, wird zumindest ein Objekt erkannt und über einen Zeitverlauf verfolgt.

[0008] Erfindungsgemäß wird das zumindest eine Objekt mittels zumindest eines Klassifizierungsverfahrens erkannt und Disparitätsbildpunkte des zumindest einen Objekts werden zu zumindest einem Cluster zusammengefasst, wobei der zumindest eine Cluster des zumindest einen erkannten Objekts durch zumindest einen vorgegebenen einfachen geometrischen Körper approximiert wird.

[0009] Durch diese Approximation des zumindest einen Clusters des zumindest einen erkannten Objekts durch den zumindest einen vorgegebenen einfachen geometrischen Körper wird auch das zumindest eine Objekt selbst durch diesen geometrischen Körper approximiert. Auf diese Weise können Objekte beliebiger Form durch ein oder mehrere einfache geometrische Körper approximiert werden, wodurch eine modellfreie Objektverfolgung ermöglicht wird.

[0010] Das Verfahren ermöglicht beispielsweise eine enge Kooperation von Menschen und Robotern, zum Beispiel in Fertigungsprozessen der Automobilindustrie. Bisher ist eine derartige Kooperation aufgrund von Sicherheitsaspekten stark eingeschränkt, da mittels Verfahren nach dem Stand der Technik ein genauer Ort, eine Körperhaltung oder ein Bewegungsverhalten des Menschen nicht mit ausreichender Sicherheit erkannt werden kann. Daher sind bisher eine Vielzahl von aufwendigen und teuren Sicherheitssystemen erforderlich, wie beispielsweise Metallzäune, Lichtschranken, Laserscanner oder kombinierte Systeme, um Menschen von einem Produktionsumfeld einer potentiell gefährlichen Maschine fernzuhalten.

[0011] Mittels des Verfahrens ist eine sichere Erkennung von Objekten und eine Verfolgung von Bewegungen der erkannten Objekte sowie eine Prädizierung ihres Bewegungsverhaltens ermöglicht. Dabei erfolgt die Objekterkennung und Objektverfolgung

modellfrei, d. h. es wird kein Modell des zu erkennenden und zu verfolgenden Objekts vorgegeben. Dies hat den Vorteil, dass im Vorfeld des Verfahrens kein explizites Modellwissen über das zu verfolgende Objekt vorhanden sein muss, wodurch sich beliebige Objekte erkennen und verfolgen lassen, indem sie über eine oder mehrere einfache geometrische Körper, beispielsweise Quader oder Ellipsoide, approximiert werden. In einer nachfolgenden Stufe kann durch eine Trajektorienklassifikation, welche aus M. Hahn, L. Krüger, C. Wähler. 3D Action Recognition and Long-term Prediction of Human Motion. In: A. Gasteratos, M. Vincze, J. Tsotsos (eds.). Proc. Int. Conf. on Computer Vision Systems, Santorini, Greece. Lecture Notes in Computer Science 5008, pp. 23–32, Springer-Verlag Berlin Heidelberg, 2008 bekannt ist, entschieden werden, ob das verfolgte Objekt beispielsweise zu einem Mensch oder zu einem Roboter gehört.

[0012] Auf diese Weise ermöglicht das Verfahren eine sichere Mensch-Roboter-Interaktion und somit einen gemeinsamen Fertigungsprozess von Mensch und Roboter, d. h. deren enge Zusammenarbeit ohne eine Sicherheitsbeeinträchtigung für die beteiligten Menschen. Dabei wird beispielsweise ein Raumbereich um den Roboter herum überwacht, zum Beispiel mittels eines Stereokamerasystems oder mittels eines multiokularen Kamerasystems, wobei mittels des Verfahrens alle sich bewegenden Objekte innerhalb des überwachten Raumbereichs erkannt und verfolgt werden. Auf diese Weise ist in Abhängigkeit von einer aktuellen Position und einem aktuellen Bewegungszustand jedem Objekt mittels der Trajektorienklassifikation ein orts- und zeitabhängig variabler Gefährdungsgrad zuweisbar.

[0013] Des Weiteren ist das Verfahren beispielsweise auch in Fahrzeugen zur Überwachung eines Verkehrsraumes um das Fahrzeug herum einsetzbar, wodurch Fahrerassistenzsysteme optimierbar sind. Hierbei ist insbesondere eine Erkennung und Verfolgung von anderen Verkehrsteilnehmern in Kreuzungsbereichen sehr wichtig, um Kollisionen zu vermeiden. Da jedoch gerade in Kreuzungsbereichen sehr viele Verkehrsteilnehmer in einem engen Verkehrsraum komplexe, durch starke Beschleunigungen oder Abbremsungen gekennzeichnete Bewegungen ausführen, ist eine sichere Erkennung und Verfolgung mittels Verfahren nach dem Stand der Technik nicht durchführbar, wohingegen mittels des erfindungsgemäßen Verfahrens und dessen Ausführungsformen auch sehr komplexe Bewegungen robust über einen Zeitverlauf hinweg verfolgt werden können, da für jeden erfassten dreidimensionalen Disparitätsbildpunkt eine Wahrscheinlichkeit ermittelt werden kann, ob dieser zum erkannten und verfolgten Objekt gehört oder nicht.

[0014] Ausführungsbeispiele der Erfindung werden

anhand von Zeichnungen näher erläutert.

[0015] Dabei zeigen:

[0016] [Fig. 1](#) drei mittels einer Kameraanordnung erfasste Bilder,

[0017] [Fig. 2](#) einen einfachen geometrischen Körper in einem dreidimensionalen Koordinatensystem, und

[0018] [Fig. 3](#) ein Referenz-Merkmalshistogramm.

[0019] Einander entsprechende Teile sind in allen Figuren mit den gleichen Bezugszeichen versehen.

[0020] Zur Durchführung des Verfahrens wird ein zu überwachender Raumbereich mittels einer Kameraanordnung überwacht, wobei die Kameraanordnung eine Mehrzahl von Kameras umfasst. Eine derartige Kameraanordnung ist beispielsweise ein Stereokamerasystem oder ein multiokulares Kamerasystem.

[0021] Der zu überwachende Raumbereich ist beispielsweise ein Umgebungsbereich eines Roboters, zum Beispiel in der Fahrzeugproduktion. Dadurch ist mittels des Verfahrens eine sichere Mensch-Roboter-Interaktion und somit ein gemeinsamer Fertigungsprozess von Mensch und Roboter ohne eine Sicherheitsbeeinträchtigung möglich.

[0022] Der zu überwachende Raumbereich kann des Weiteren beispielsweise auch eine Umgebung eines Fahrzeugs sein, welche mittels eines am Fahrzeug angeordneten und mit Fahrerassistenzsystemen des Fahrzeugs gekoppelten Stereokamerasystems überwacht wird. Hierbei ist insbesondere eine Erkennung und Verfolgung von anderen Verkehrsteilnehmern in Kreuzungsbereichen sehr wichtig, um Kollisionen zu vermeiden. Da jedoch gerade in Kreuzungsbereichen sehr viele Verkehrsteilnehmer in einem engen Verkehrsraum komplexe, durch starke Beschleunigungen oder Abbremsungen gekennzeichnete Bewegungen ausführen, ist eine sichere Erkennung und Verfolgung mittels Verfahren nach dem Stand der Technik nicht durchführbar, wohingegen mittels des erfindungsgemäßen Verfahrens und dessen Ausführungsformen auch sehr komplexe Bewegungen robust über einen Zeitverlauf hinweg verfolgt werden können.

[0023] Im in [Fig. 1](#) dargestellten Ausführungsbeispiel sind drei Bilder B1, B2, B3 dargestellt, welche mittels der Kameraanordnung erfasst wurden, wobei die hier nicht näher dargestellte Kameraanordnung zur Erfassung dieser Bilder B1, B2, B3 ein multiokulares Kamerasystem ist, welches aus drei Kameras zur Erfassung von Bilddaten des zu überwachenden Raumbereichs gebildet ist. Die Bilder B1, B2, B3 werden anschließend einer nicht näher dargestellten

Verarbeitungseinheit zur multiokularen Bildverarbeitung zugeführt. Bei dieser multiokularen Bildverarbeitung werden Korrespondenzen von Bildpunkten (auch Pixel genannt) in den multiokular aufgenommenen Bildern B1, B2, B3 bestimmt.

[0024] Hierbei werden die Bilder B1, B2, B3, wie beispielsweise in U. Franke, A. Joos. Real-time stereo vision for urban traffic scene understanding. In Procs. IEEE Intelligent Vehicles Symposium, pp. 273–278, Dearborn, USA, 2000 beschrieben, in einem Algorithmus derart verarbeitet, dass Koordinaten eines Bildpunkts des einen Bilds B1 mit Koordinaten eines als potentiell korrespondierend betrachteten Bildpunkts des zweiten Bilds B2 und Koordinaten eines als potentiell korrespondierend betrachteten Bildpunkts des dritten Bilds B3 verglichen werden. Aus einem Abstand der Bildpunkte zueinander, der so genannten Disparität, und einem bekannten Abstand der Kameras wird der Abstand eines Objekts O, welches die Bildpunkte aufweist, zu den Kameras bestimmt. Nach diesem Algorithmus werden Disparitäten für alle Bildpunkte der Bilder B1, B2, B3 ermittelt und ein Disparitätsbild oder eine Disparitätskarte erzeugt, welche mittels Disparitätsbildpunkten eine dreidimensionale Repräsentation aller erfassten Objekte O in ihrem Kontext darstellen. Auf diese Weise ist die Entfernung und räumliche Lage der erfassten Objekte O im Verhältnis zu den Kameras ermittelbar.

[0025] Durch eine Analyse zeitlich aufeinander folgender Bildkombinationen der drei Kameras ist eine Bewegung der erfassten Bildpunkte zwischen einem ersten Zeitpunkt und einem zweiten Zeitpunkt durch eine Ermittlung eines optischen Flusses der Bildpunkte, wie in A. Wedel, T. Pock, C. Zach, H. Bischof, D. Cremers. An improved algorithm for tv-l1 optical flow computation. In Proceedings of the Dagstuhl Visual Motion Analysis Workshop, 2008 beschrieben, ermittelbar. Aus diesen Informationen ist ein so genanntes dreidimensionales Bewegungsfeld, auch Scene-Flow genannt, d. h. eine Abstandsänderung der Bildpunkte zu den Kameras ermittelbar. Dadurch können alle Disparitätsbildpunkte ermittelt werden, welche sich in den aufeinander folgenden Bildkombinationen bewegt haben.

[0026] Alle sich bewegenden Disparitätsbildpunkte werden mittels zumindest eines Klassifizierungsverfahrens, vorzugsweise mittels mehrerer stufenweise nacheinander durchgeführter Klassifizierungsverfahren analysiert, um zu ermitteln, welche Disparitätsbildpunkte zu einem sich bewegenden und zu verfolgenden Objekt O gehören. Diese Disparitätsbildpunkte werden zu einem so genannten Cluster zusammengefasst.

[0027] Dabei ist ein erstes Klassifizierungsverfahren beispielsweise ein graphenbasiertes Klassifizierungsverfahren, wie es in H. H. Bock. Automatische

Klassifikation. Vandenhoeck & Ruprecht, Göttingen, 1974 beschrieben wird, wodurch relativ große Cluster erzielt werden. In [Fig. 1](#) wird von den Kameras ein Mensch erfasst, wobei sich im Zeitverlauf beispielsweise ein Arm A inklusive einer Hand bewegt. Daher werden mittels des ersten Klassifizierungsverfahrens alle Disparitätsbildpunkte, welche sich bewegen, d. h. alle Disparitätsbildpunkte des Arms A und der Hand zu einem Cluster zusammengefasst.

[0028] In einem zweiten Klassifizierungsverfahren, beispielsweise mittels des so genannten Mean-Shift-Clustering, wie in Y. Cheng. Mean Shift, Mode Seeking, and Clustering, IEEE Trans. Pattern Analysis Machine Intell., Vol. 17, No. 8, pp. 790–799, 1995 beschrieben, wird diese Klassifizierung verfeinert, so dass Disparitätsbildpunkte mit einem ähnlichen Ort und einer ähnlichen zugehörigen Geschwindigkeit zu einem Cluster zusammengefasst werden und auf diese Weise das eigentlich zu verfolgende Objekt O, im hier dargestellten Beispiel die Hand, erkannt wird. D. h. ein solches verfeinertes Cluster umfasst im hier dargestellten Ausführungsbeispiel nur die Hand. Diese Cluster, im hier dargestellten Beispiel das Cluster, welches die Disparitätsbildpunkte der Hand umfasst, werden zur Objektverfolgung durch vorgegebene einfache geometrische Körper K, beispielsweise Quader oder Ellipsoide, approximiert.

[0029] Im hier dargestellten Ausführungsbeispiel wird, wie dargestellt, die Hand durch einen geometrischen Körper K approximiert, welcher als Ellipsoid ausgebildet ist. Dieser geometrische Körper K ist in [Fig. 2](#) in einem dreidimensionalen Weltkoordinatensystem dargestellt, wodurch eine dreidimensionale Pose des geometrischen Körpers K und somit des durch diesen approximierten zu verfolgenden Objekts O ermittelbar ist, wobei mit der dreidimensionalen Pose sowohl eine Position des geometrischen Körpers K und somit des zu verfolgenden Objekts O im zu überwachenden Raumbereich, ermittelbar durch dreidimensionale Weltkoordinaten x , y , z der Disparitätsbildpunkte, als auch dessen Ausrichtung gemeint ist, welche anhand eines Drehwinkels ϕ ermittelbar ist.

[0030] Auf diese Weise sind Objekte O beliebiger Form durch ein oder mehrere dieser vorgegebenen einfachen geometrischen Körper K approximierbar, wodurch, im Gegensatz zum Stand der Technik, eine modellfreie Objektverfolgung ermöglicht ist, d. h. zur Erkennung und Verfolgung der Objekte O ist kein vorgegebenes Modell dieser Objekte O erforderlich, bei welchem Objekte O, welche nicht diesem vorgegebenen Modell entsprechen, nicht erkannt werden würden und somit nicht verfolgt werden könnten.

[0031] Zur Objektverfolgung werden die aus D. Comaniciu, V. Ramesh, P. Meer: Real-Time Tracking of Non-Rigid Objects using Mean Shift, IEEE Conf.

Computer Vision and Pattern Recognition (CV-PR'00), Hilton Head Island, South Carolina, Vol. 2, pp. 142–149, 2000, sowie aus D. Comaniciu, V. Ramesh, P. Meer: Kernel-Based Object Tracking, IEEE Trans. Pattern Analysis Machine Intell., Vol. 25, No. 5, pp. 564–575, 2003 und aus G. R. Gradski, Real time face and object tracking as a component of a perceptual user interface. In WACV'98: Proceedings of the 4th IEEE Workshop on Applications of Computer Vision, pp. 214–219, Washington, DC, USA. IEEE Computer Society bekannten Mean-Shift-Tracking Verfahren dreidimensional erweitert. Dabei wird für jedes erkannte dreidimensionale Objekt O ein Referenzmodell, ein so genanntes Referenz-Merkmalshistogramm RM ermittelt. Dieses Referenz-Merkmalshistogramm RM stellt eine Verteilung relativer Häufigkeiten h von Bildpunktwerten BW des erfassten und zu verfolgenden Objekts O dar. Diese Bildpunktwerte BW sind beispielsweise Graustufenwerte oder Farbwerte.

[0032] Ein solches Referenz-Merkmalshistogramm RM ist in [Fig. 3](#) dargestellt. Dabei wird angenommen, dass alle Disparitätsbildpunkte des zu verfolgenden Objekts O auf dem approximierten geometrischen Körper K die gleiche Tiefe haben, d. h. dass der zu verfolgende geometrische Körper K flach ist. Auf diese Weise ist jeder sichtbare Disparitätsbildpunkt auf einer Oberfläche des einfachen geometrischen Körpers K sehr einfach aus dem dreidimensionalen Koordinatensystem in Koordinatensysteme der einzelnen Kameras projizierbar.

[0033] Das Referenz-Merkmalshistogramm RM repräsentiert Wahrscheinlichkeitswerte, die von der Häufigkeit im Referenz-Merkmalshistogramm RM abhängen und aussagen, wie wahrscheinlich es ist, dass ein Disparitätsbildpunkt mit den jeweiligen Bildpunktwerten BW in den verwendeten Kameras zum zu verfolgenden Objekt O gehört. Die Zuordnung der Bildpunktwerte BW erfolgt über eine Berechnungsvorschrift, welche die Bildpunktwerte BW aus allen Bildern B1, B2, B3 ermittelt und einem jeweiligen so genannten Histogramm-Bin HB1 bis HB5 zuordnet, zum Beispiel über den Mittelwert der Bildpunktwerte BW. Dabei umfasst jedes der dargestellten Histogramm-Bin HB1 bis HB5 einen Wertebereich verschiedener Bildpunktwerte BW mit jeweils einer unteren und einer oberen Bildpunktwertgrenze. Aufgrund dieser Zuordnung ergeben sich im Referenz-Merkmalshistogramm RM verschiedene relative Häufigkeiten h für verschiedene Bildpunktwerte BW.

[0034] Dieses Referenz-Merkmalshistogramm RM wird verwendet, um das zu verfolgende Objekt O in Bildern B1, B2, B3, welche in nachfolgenden Zeitschritten erfasst werden, wieder aufzufinden. Ein Suchrasterbereich, dessen Bildpunkte mit dem Referenz-Merkmalshistogramm RM verglichen werden, ist dabei in der dreidimensionalen Erweiterung der

Mean-Shift-Verfahren nicht durch Bildkoordinaten, sondern durch die dreidimensionalen Weltkoordinaten x, y, z definiert, wobei der Suchrasterbereich ein dreidimensionaler Bereich ist, welcher parallel zu einer XY-Ebene des jeweiligen Kamerakoordinatensystems verläuft. Dadurch besitzt jeder Bildpunkt des Suchrasterbereichs die gleiche Tiefe.

[0035] Jedem Bildpunkt innerhalb des Suchrasterbereichs wird mittels des Referenz-Merkmalshistogramms RM und einer jeweiligen Berechnungsvorschrift eine Wahrscheinlichkeit in einem Intervall zwischen 0 und 1 zugewiesen, ob der Bildpunkt zum zu verfolgenden Objekt O gehört oder nicht. Dabei wird das Referenz-Merkmalshistogramm RM als Datenbank, d. h. als so genannter Lookup-Table verwendet, wobei mit Hilfe der Berechnungsvorschrift auf die Histogramm-Bins HB des Referenz-Merkmalshistogramms RM zugegriffen wird. Der Suchrasterbereich wird dabei in Abhängigkeit von der in einem vorhergehenden Zeitschritt erfassten Position des zu verfolgenden Objekts O festgelegt. Eine Größe des Suchrasterbereichs ist abhängig von einer Auflösung der Bilder B1, B2, B3, welche wiederum abhängig von der Entfernung des zu verfolgenden Objekts O sind.

[0036] Der geometrische Körper K, durch welchen das zu verfolgende Objekt O approximiert wurde, wird nun mit dem in D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603–619, 2002 beschriebenen Mean-Shift-Algorithmus an den Suchrasterbereich angepasst, indem innerhalb des Suchrasterbereichs eine Region gesucht wird, in welcher die Bildpunkte die höchste Wahrscheinlichkeit aufweisen, zu dem zu verfolgenden Objekt O zu gehören.

[0037] Da der Suchrasterbereich flach ist, d. h. zweidimensional ist und daher keine Ausdehnung in die Tiefe aufweist, wird dies mit einer entsprechenden abgeflachten, d. h. zweidimensionalen Variante des geometrischen Körpers K durchgeführt, d. h. ist der geometrische Körper K ein Quader, so ist die zweidimensionale Variante ein Rechteck, ist der geometrische Körper K ein Ellipsoid, so ist die zweidimensionale Variante eine Ellipse. Ist das zu verfolgende Objekt O auf diese Weise in den in diesem Zeitschritt erfassten Bildern B1, B2, B3 aufgefunden worden, wird ein Kandidaten-Merkmalshistogramm des in diesen Bildern B1, B2, B3 ermittelten Objekts O erstellt und mit dem Referenz-Merkmalshistogramm RM verglichen, um festzustellen, ob es sich tatsächlich um das zu verfolgende Objekt O handelt.

[0038] In einem zweiten Schritt wird die Entfernung des Objekts O geschätzt. Dazu wird der Mean-Shift-Algorithmus nach D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature

Space Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603–619, 2002 für alle Disparitätsbildpunkte des dreidimensionalen Bewegungsfeldes durchgeführt, wobei die Disparitätsbildpunkte sowohl mittels eines Kernels als auch mit dem Referenz-Merkmalshistogramm RM gewichtet werden. Bei einer Gewichtung der Disparitätsbildpunkte mittels des Kernels erhalten die Disparitätsbildpunkte, welche einem Mittelpunkt des geometrischen Körpers K , d. h. im hier dargestellten Ausführungsbeispiel dem Mittelpunkt des Ellipsoids oder in einem anderen Beispiel dem Mittelpunkt eines Rechtecks am nächsten liegen, eine höhere Wahrscheinlichkeit als entferntere Disparitätsbildpunkte. Dadurch wird eine Robustheit der Objektverfolgung erhöht, wodurch ein zeitlich stabiler Verfolgungsprozess erreicht wird. Auf diese Weise sind alle Disparitätsbildpunkte ermittelbar, welche zum zu verfolgenden Objekt O gehören, und somit auch dessen aktuelle Position.

[0039] Durch einen Vergleich mit der Position des zu verfolgenden Objekts O im vorherigen Zeitschritt ist somit eine Bewegungstrajektorie des zu verfolgenden Objekts O ermittelbar. In einer nachfolgenden Stufe kann durch eine Trajektorienklassifikation, welche aus M. Hahn, L. Krüger, C. Wähler. 3D Action Recognition and Long-term Prediction of Human Motion. In: A. Gasteratos, M. Vincze, J. Tsotsos (eds.). Proc. Int. Conf. an Computer Vision Systems, Santorini, Greece. Lecture Notes in Computer Science 5008, pp. 23–32, Springer-Verlag Berlin Heidelberg, 2008 bekannt ist, beispielsweise entschieden werden, ob das verfolgte Objekt O zu einem Mensch oder zu einem Roboter gehört, wenn das Verfahren zur Überwachung eines Roboterarbeitsplatzes eingesetzt wird, oder welche Art von Verkehrsteilnehmer als Objekt O erkannt und verfolgt wird, wenn das Kamerasystem zur Überwachung eines Umfeldbereichs eines Fahrzeugs eingesetzt wird.

[0040] Durch den Vergleich des Referenz-Merkmalshistogramms RM mit dem Kandidaten-Merkmalshistogramm kann außerdem festgestellt werden, wie ähnlich das Aussehen des Objekts O an der aktuell ermittelten dreidimensionalen Pose zu dem vorherigen Aussehen ist. Auf diese Weise wird erkannt, wenn das zu verfolgende Objekt O verloren gegangen ist, d. h. nicht mehr erfasst und verfolgt wird, beispielsweise wenn es sich aus dem zu überwachenden Raumbereich entfernt hat. Sollte sich das zu verfolgende Objekt O weiterhin in dem zu überwachenden Raumbereich befinden oder erneut in diesen eintreten, ist es durch einen Neustart des Verfahrens und mittels des beschriebenen Verfahrensablaufs erneut erfassbar. Daher ist es sinnvoll, dass das Verfahren nach jedem Verlust des zu verfolgenden Objekts O sofort erneut startet.

[0041] Insbesondere das Referenz-Merkmalshisto-

gramm RM ist auch bei anderen Objektverfolgungsalgorithmen, beispielsweise bei den in M. Hahn, L. Krüger, C. Wähler, H.-M. Groß. Tracking of Human Body Parts using the Multiocular Contracting Curve Density Algorithm. Proc. Int. Conf. an 3-D Digital Imaging and Modeling, pp. 257–264, Montreal, Canada, 2007 und M. Hahn, L. Krüger, C. Wähler. Spatio-Temporal 3D Pose Estimation and Tracking of Human Body Parts Using the Shape Flow Algorithm. Proc. Int. Conf. an Pattern Recognition, Tampa, USA, 2008 beschriebenen Objektverfolgungsalgorithmen verwendbar, um einen Verlust der Objektverfolgung zu erkennen, wobei mit dem beschriebenen Verfahren das zu verfolgende Objekt O wieder erkannt werden kann.

[0042] Des Weiteren sind mittels des Referenz-Merkmalshistogramms RM auch in anderen Anwendungsbereichen dreidimensionale Bildpunktewolken untersuchbar, um zu erkennen, ob die dreidimensionale Bildpunktewolke oder Teile davon zu einem durch das Referenz-Merkmalshistogramm RM repräsentierten Objekt O gehören.

Bezugszeichenliste

A	Arm
B1, B2, B3	Bilder
BW	Bildpunktewerte
h	relative Häufigkeit
HB1 bis HB5	Histogramm-Bin
K	Körper
O	Objekt
RM	Referenz-Merkmalshistogramm
x, y, z	Weltkoordinaten
ϕ	Drehwinkel

ZITATE ENTHALTEN IN DER BESCHREIBUNG

Diese Liste der vom Anmelder aufgeführten Dokumente wurde automatisiert erzeugt und ist ausschließlich zur besseren Information des Lesers aufgenommen. Die Liste ist nicht Bestandteil der deutschen Patent- bzw. Gebrauchsmusteranmeldung. Das DPMA übernimmt keinerlei Haftung für etwaige Fehler oder Auslassungen.

Zitierte Patentliteratur

- US 2003/0081815 A1 [\[0002\]](#)

Zitierte Nicht-Patentliteratur

- D. Comaniciu, V. Ramesh, P. Meer: Real-Time Tracking of Non-Rigid Objects using Mean Shift, IEEE Conf. Computer Vision and Pattern Recognition (CVPR'00), Hilton Head Island, South Carolina, Vol. 2, pp. 142–149, 2000 [\[0003\]](#)
- D. Comaniciu, V. Ramesh, P. Meer: Kernel-Based Object Tracking, IEEE Trans. Pattern Analysis Machine Intell., Vol. 25, No. 5, pp. 564–575, 2003 [\[0003\]](#)
- G. R. Gradski, Real time face and object tracking as a component of a perceptual user interface. In WACV'98: Proceedings of the 4th IEEE Workshop an Applications of Computer Vision, pp. 214–219, Washington, DC, USA. IEEE Computer Society [\[0003\]](#)
- M. Hahn, L. Krüger, C. Wähler. 3D Action Recognition and Long-term Prediction of Human Motion. In: A. Gasteratos, M. Vincze, J. Tsotsos (eds.). Proc. Int. Conf. an Computer Vision Systems, Santorini, Greece. Lecture Notes in Computer Science 5008, pp. 23–32, Springer-Verlag Berlin Heidelberg, 2008 [\[0011\]](#)
- U. Franke, A. Joos. Real-time stereo vision for urban traffic scene understanding. In Procs. IEEE Intelligent Vehicles Symposium, pp. 273–278, Dearborn, USA, 2000 [\[0024\]](#)
- A. Wedel, T. Pock, C. Zach, H. Bischof, D. Cremers. An improved algorithm for tv-l1 optical flow computation. In Proceedings of the Dagstuhl Visual Motion Analysis Workshop, 2008 [\[0025\]](#)
- H. H. Bock. Automatische Klassifikation. Vandenhoeck & Ruprecht, Göttingen, 1974 [\[0027\]](#)
- Y. Cheng. Mean Shift, Mode Seeking, and Clustering, IEEE Trans. Pattern Analysis Machine Intell., Vol. 17, No. 8, pp. 790–799, 1995 [\[0028\]](#)
- D. Comaniciu, V. Ramesh, P. Meer: Real-Time Tracking of Non-Rigid Objects using Mean Shift, IEEE Conf. Computer Vision and Pattern Recognition (CVPR'00), Hilton Head Island, South Carolina, Vol. 2, pp. 142–149, 2000 [\[0031\]](#)
- D. Comaniciu, V. Ramesh, P. Meer: Kernel-Based Object Tracking, IEEE Trans. Pattern Analysis Machine Intell., Vol. 25, No. 5, pp. 564–575, 2003 [\[0031\]](#)

- G. R. Gradski, Real time face and object tracking as a component of a perceptual user interface. In WACV'98: Proceedings of the 4th IEEE Workshop an Applications of Computer Vision, pp. 214–219, Washington, DC, USA [\[0031\]](#)
- D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603–619, 2002 [\[0036\]](#)
- D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603–619, 2002 [\[0038\]](#)
- M. Hahn, L. Krüger, C. Wähler. 3D Action Recognition and Long-term Prediction of Human Motion. In: A. Gasteratos, M. Vincze, J. Tsotsos (eds.). Proc. Int. Conf. an Computer Vision Systems, Santorini, Greece. Lecture Notes in Computer Science 5008, pp. 23–32, Springer-Verlag Berlin Heidelberg, 2008 [\[0039\]](#)
- M. Hahn, L. Krüger, C. Wähler, H.-M. Groß. Tracking of Human Body Parts using the Multiocular Contracting Curve Density Algorithm. Proc. Int. Conf. an 3-D Digital Imaging and Modeling, pp. 257–264, Montreal, Canada, 2007 [\[0041\]](#)
- M. Hahn, L. Krüger, C. Wähler. Spatio-Temporal 3D Pose Estimation and Tracking of Human Body Parts Using the Shape Flow Algorithm. Proc. Int. Conf. an Pattern Recognition, Tampa, USA, 2008 [\[0041\]](#)

Patentansprüche

1. Verfahren zur Objekterkennung und Objektverfolgung unter Verwendung von mittels einer Kameraanordnung erfasster Bilder (B1, B2, B3), wobei die Kameraanordnung eine Mehrzahl von Kameras umfasst, wobei mittels einer Verarbeitungseinheit aus den Bildern (B1, B2, B3) der Kamera ein dreidimensionales Bewegungsfeld erstellt wird, wobei zumindest ein Objekt (O) erkannt und über einen Zeitverlauf verfolgt wird, **dadurch gekennzeichnet**, dass das zumindest eine Objekt (O) mittels zumindest eines Klassifizierungsverfahrens erkannt und Disparitätsbildpunkte des zumindest einen Objekts (O) zu zumindest einem Cluster zusammengefasst werden, wobei der zumindest eine Cluster des zumindest einen erkannten Objekts (O) durch zumindest einen vorgegebenen einfachen geometrischen Körper (K) approximiert wird.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass zur Objekterkennung nacheinander ein erstes Klassifizierungsverfahren und ein zweites, verfeinertes Klassifizierungsverfahren durchgeführt wird.

3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, dass ein Referenz-Merkmalshistogramm (RM) des erkannten Objekts (O) ermittelt wird.

4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, dass eine Bewegungstrajektorie des zumindest einen Objekts (O) ermittelt wird.

5. Verfahren nach Anspruch 4, dadurch gekennzeichnet, dass die ermittelte Bewegungstrajektorie klassifiziert wird und anhand dieser Klassifizierung das Objekt (O) identifiziert wird.

6. Verwendung des Verfahrens nach einem der Ansprüche 1 bis 5 zur Überwachung eines Roboterarbeitsplatzes.

7. Verwendung des Verfahrens nach einem der Ansprüche 1 bis 5 in einem Fahrzeug zur Überwachung eines Umfeldbereichs des Fahrzeugs.

Es folgt ein Blatt Zeichnungen

Anhängende Zeichnungen

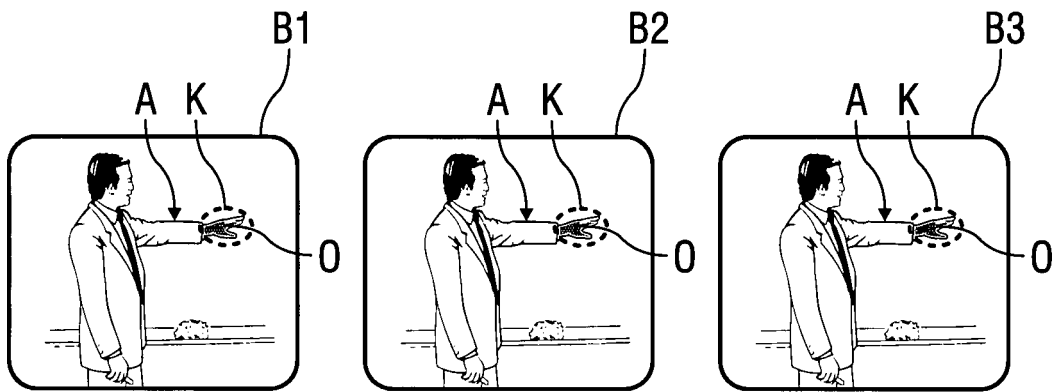


FIG 1

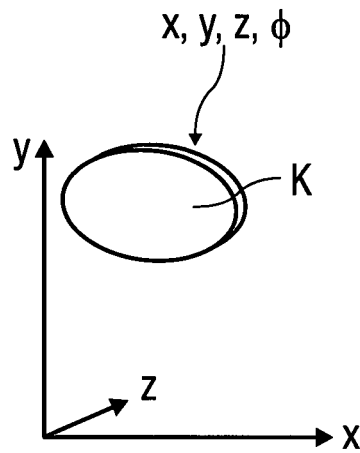


FIG 2

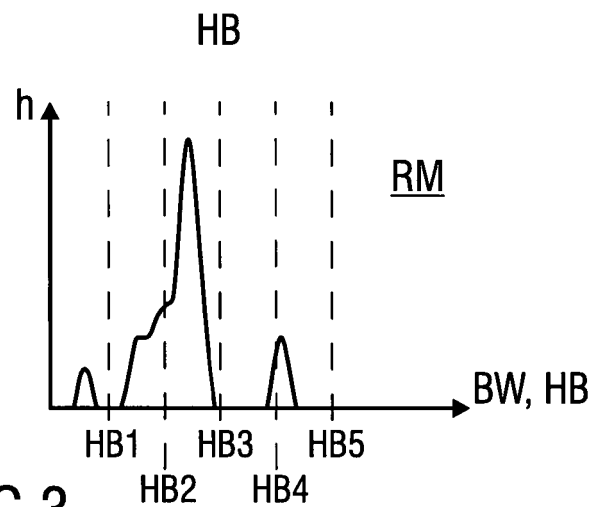


FIG 3