

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

(19) Weltorganisation für geistiges Eigentum

Internationales Büro

(43) Internationales Veröffentlichungsdatum  
25. August 2016 (25.08.2016)



(10) Internationale Veröffentlichungsnummer  
**WO 2016/131812 A1**

- (51) Internationale Patentklassifikation:  
**G06K 9/00** (2006.01) **G06K 9/32** (2006.01)
- (21) Internationales Aktenzeichen: PCT/EP2016/053242
- (22) Internationales Anmeldedatum:  
16. Februar 2016 (16.02.2016)
- (25) Einreichungssprache: Deutsch
- (26) Veröffentlichungssprache: Deutsch
- (30) Angaben zur Priorität:  
102015102369.5 19. Februar 2015 (19.02.2015) DE
- (71) Anmelder: **BUNDESDRUCKEREI GMBH** [DE/DE];  
Kommandantenstr. 18, 10969 Berlin (DE).
- (72) Erfinder: **HARTL, Andreas**; Hüttenbrennergasse 17/P/4,  
8010 Graz (AT). **SCHMALSTIEG, Dieter**; Ursprungweg  
146, 8045 Graz (AT). **DRESSEL, Olaf**; Straße der  
Gemeinschaft 25B, 14641 Wustermark (DE).
- (74) Anwalt: **KLINSKI, Robert**; Elsenheimerstraße 65, 80687  
München (DE).
- (81) Bestimmungsstaaten (soweit nicht anders angegeben, für  
jede verfügbare nationale Schutzrechtsart): AE, AG, AL,

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

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

Veröffentlicht:

— mit internationalem Recherchenbericht (Artikel 21 Absatz  
3)

(54) Title: MOBILE DEVICE FOR CAPTURING A TEXT RANGE ON AN IDENTIFICATION DOCUMENT

(54) Bezeichnung : MOBILGERÄT ZUM ERFASSEN EINES TEXTBEREICHES AUF EINEM IDENTIFIKATIONSDOKUMENT

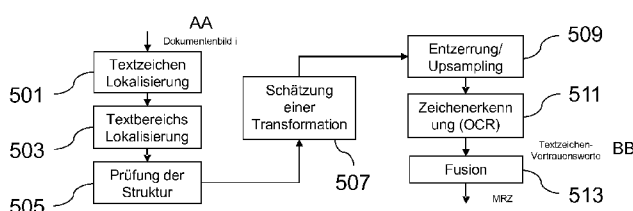


Fig. 5

AA Document image i  
BB Text character confidence values  
501 Text character location  
503 Text range location  
505 Check on the structure  
507 Estimate of a transformation  
509 Equalisation/up-sampling  
511 Character recognition (OCR)

(57) Abstract: The invention relates to a mobile device (100) for capturing a text range on an identification document, wherein the text range has a plurality of text characters in a predetermined arrangement in accordance with a predetermined arrangement extent, having a photographic camera (101), which is designed to capture an image of the identification document in order to obtain a document image, and a processor (103), which is designed to segment the document image in order to obtain a plurality of image segments, to select a plurality of text character image segments from the plurality of image segments, wherein the text character image segments each represent a text character, to determine a plurality of text character groups on the basis of the plurality of text character image segments, wherein the text character groups each comprise a series of text character image segments, and to compare a plurality of arrangement extents for the plurality of text character groups with the predetermined arrangement extent in order to capture the text range on the identification document.

(57) Zusammenfassung:

[Fortsetzung auf der nächsten Seite]



---

Die Erfindung betrifft ein Mobilgerät (100) zum Erfassen eines Textbereiches auf einem Identifikationsdokument, wobei der Textbereich eine Mehrzahl von Textzeichen in einer vorbestimmten Anordnung gemäß einem vorbestimmten Anordnungsmaß aufweist, mit einer Bildkamera (101), welche ausgebildet ist, ein Bild des Identifikationsdokumentes zu erfassen, um ein Dokumentenbild zu erhalten, und einem Prozessor (103), welcher ausgebildet ist, das Dokumentenbild zu segmentieren, um eine Mehrzahl von Bildsegmenten zu erhalten, eine Mehrzahl von Textzeichenbildsegmenten aus der Mehrzahl von Bildsegmenten auszuwählen, wobei die Textzeichenbildsegmente jeweils ein Textzeichen repräsentieren, eine Mehrzahl von Textzeichengruppen auf Basis der Mehrzahl von Textzeichenbildsegmenten zu bestimmen, wobei die Textzeichengruppen jeweils eine Folge von Textzeichenbildsegmenten umfassen, und eine Mehrzahl von Anordnungsmaßen der Mehrzahl von Textzeichengruppen mit dem vorbestimmten Anordnungsmaß zu vergleichen, um den Textbereich auf dem Identifikationsdokument zu erfassen.

## 5 Mobilgerät zum Erfassen eines Textbereiches auf einem Identifikationsdokument

Die vorliegende Erfindung betrifft das Gebiet der Erfassung von Textbereichen auf Identifikationsdokumenten.

10 Die Erfassung von Textbereichen, beispielsweise maschinen-lesbarer Zonen, auf Identifikationsdokumenten ist für die Gültigkeitsprüfung der Identifikationsdokumente oder für die Identifizierung von Personen von besonderer Bedeutung. Insbesondere bei Grenzkontrollen ist eine effiziente Erfassung mit hohen Genauigkeiten wünschenswert.

15 Üblicherweise erfolgt eine Erfassung von Textbereichen unter Verwendung stationärer Erfassungsgeräte, auf welche die Identifikationsdokumente manuell aufgelegt werden. Ferner können Mobilgeräte für die Erfassung von Textbereichen eingesetzt werden, wobei eine manuelle orthogonale Ausrichtung der Identifikationsdokumente bezüglich der Mobilgeräte durchgeführt wird. Sowohl bei stationären Erfassungsgeräten als auch bei  
20 Mobilgeräten werden folglich manuelle Handlungen durch einen Benutzer durchgeführt, welche die Erfassungszeit erhöhen und mit Effizienzverlusten verbunden sind.

In V. Bessmeltsev, E. Bulushev, N. Goloshevsky, "High-speed OCR algorithm for portable passport readers", GraphiCon, 2011, wird eine optische Zeichenerkennung für tragbare  
25 Reisepass-Lesegeräte beschrieben.

In Y. Wu, L. Xue, C. Li, "TDARS, a Fusion Based AR System for Machine Readable Travel Documents", HCI International, 2007, wird ein fusionsbasiertes Augmented Reality (AR) System für maschinen-lesbare Reisedokumente beschrieben.  
30

Es ist eine Aufgabe der vorliegenden Erfindung, ein effizientes Konzept zum Erfassen eines Textbereiches auf einem Identifikationsdokument zu schaffen.

Diese Aufgabe wird durch die Merkmale der unabhängigen Ansprüche gelöst. Vorteilhafte  
35 Weiterbildungsformen sind Gegenstand der abhängigen Patentansprüche, der Beschreibung sowie der Figuren.

Die Erfindung basiert auf der Erkenntnis, dass die obige Aufgabe durch ein Mobilgerät gelöst werden kann, welches eine Bildkamera und einen Prozessor umfasst, wobei die  
40 Bildkamera ausgebildet ist, ein Bild des Identifikationsdokumentes zu erfassen, um ein

5 Dokumentenbild zu erhalten, und wobei der Prozessor ausgebildet ist, den Textbereich in dem Dokumentenbild auf Basis einer Segmentierung des Dokumentenbildes zu erfassen.

Die Bildsegmente können dabei zunächst hinsichtlich typischer Eigenschaften von Textzeichen gefiltert werden, um eine Mehrzahl von Textzeichenbildsegmenten zu erhalten. Anschließend kann die Mehrzahl von Textzeichenbildsegmenten gruppiert werden, um eine Mehrzahl von Textzeichengruppen zu erhalten, welche Wörter oder Folgen von Textzeichen repräsentieren. Die Mehrzahl von Textzeichengruppen weist dabei beispielsweise jeweils ein Anordnungsmaß auf, welches mit einem vorbestimmten Anordnungsmaß des zu erfassenden Textbereiches verglichen werden kann.

15

Folglich kann ein vorbestimmtes Anordnungsmaß des Textbereiches, wie beispielsweise eine vorbestimmte Anzahl von Textzeichen je Textzeile, zur Erfassung des Textbereiches herangezogen werden. Unter Verwendung einer optischen Zeichenerkennung innerhalb des erfassten Textbereiches kann ferner eine Mehrzahl von Textzeichen des Textbereiches erfasst werden. Die optische Zeichenerkennung kann dabei einen vorbestimmten Zeichensatz zulässiger Buchstaben, Ziffern oder Symbole der Textzeichen verwenden, wodurch eine effiziente optische Zeichenerkennung bereitgestellt wird.

Der Textbereich kann eine maschinen-lesbare Zone (Machine Readable Zone, MRZ) des Identifikationsdokumentes sein, wobei die Anordnung der Textzeichen der maschinen-lesbaren Zone gemäß Standards der ICAO (International Civil Aviation Organization) standardisiert sein kann. Das vorbestimmte Anordnungsmaß, beispielsweise die Anzahl von Textzeichen je Textzeile, kann aus den Standards vorbekannt sein.

Das Konzept ermöglicht eine effiziente Erfassung des Textbereiches unmittelbar in dem Dokumentenbild, wobei auf eine vorherige Erfassung des Identifikationsdokumentes in dem Dokumentbild verzichtet werden kann. Ferner kann eine effiziente Erfassung des Textbereiches unabhängig von der Raumlage des Mobilgerätes relativ zu dem Identifikationsdokument erreicht werden, wodurch auf eine manuelle orthogonale Ausrichtung verzichtet werden kann.

Gemäß einem ersten Aspekt betrifft die Erfindung ein Mobilgerät zum Erfassen eines Textbereiches auf einem Identifikationsdokument, wobei der Textbereich eine Mehrzahl von Textzeichen in einer vorbestimmten Anordnung gemäß einem vorbestimmten Anordnungsmaß aufweist, mit einer Bildkamera, welche ausgebildet ist, ein Bild des

40

5 Identifikationsdokumentes zu erfassen, um ein Dokumentenbild zu erhalten, und einem  
Prozessor, welcher ausgebildet ist, das Dokumentenbild zu segmentieren, um eine  
Mehrzahl von Bildsegmenten zu erhalten, eine Mehrzahl von Textzeichenbildsegmenten  
aus der Mehrzahl von Bildsegmenten auszuwählen, wobei die Textzeichenbildsegmente  
10 der Mehrzahl von Textzeichenbildsegmenten zu bestimmen, wobei die  
Textzeichengruppen jeweils eine Folge von Textzeichenbildsegmenten umfassen, und  
eine Mehrzahl von Anordnungsmaßen der Mehrzahl von Textzeichengruppen mit dem  
vorbestimmten Anordnungsmaß zu vergleichen, um den Textbereich auf dem  
Identifikationsdokument zu erfassen. Dadurch wird der Vorteil erreicht, dass ein effizientes  
15 Konzept zum Erfassen eines Textbereiches auf einem Identifikationsdokument realisiert  
werden kann.

Das Mobilgerät kann ein Smartphone sein. Das vorbestimmte Anordnungsmaß kann eine  
Anzahl von Textzeichen je Textzeile anzeigen. Das vorbestimmte Anordnungsmaß kann  
20 ferner einen horizontalen und/oder vertikalen Abstand der Textzeichen anzeigen. Ein  
Textzeichen kann durch einen Buchstaben, eine Ziffer oder ein Symbol gebildet sein.

Das Identifikationsdokument kann eines der folgenden Identifikationsdokumente sein:  
Identitätsdokument, wie Personalausweis, Reisepass, Zugangskontrollausweis,  
25 Berechtigungsausweis, Unternehmensausweis, Steuerzeichen oder Ticket,  
Geburtsurkunde, Führerschein, Kraftfahrzeugausweis, Zahlungsmittel, beispielsweise  
eine Bankkarte oder Kreditkarte. Das Identifikationsdokument kann ferner einen  
elektronisch auslesbaren Schaltkreis, beispielsweise einen RFID-Chip umfassen. Das  
Identifikationsdokument kann ein- oder mehrlagig bzw. papier- und/oder kunststoffbasiert  
30 sein. Das Identifikationsdokument kann aus kunststoffbasierten Folien aufgebaut sein,  
welche zu einem Kartenkörper mittels Verkleben und/oder Laminieren zusammengefügt  
werden, wobei die Folien bevorzugt ähnliche stoffliche Eigenschaften aufweisen.

Gemäß einer Ausführungsform ist der Prozessor ausgebildet, ein Grauwertbild des  
35 Dokumentenbildes zu bestimmen. Dadurch wird der Vorteil erreicht, dass die  
Segmentierung des Dokumentenbildes effizient durchgeführt werden kann. Das  
Grauwertbild kann das Dokumentenbild zur Weiterverarbeitung bilden.

Gemäß einer Ausführungsform ist der Prozessor ausgebildet, eine Bildauflösung des  
40 Dokumentenbildes auf eine vorbestimmte Bildauflösung anzupassen. Dadurch wird der

5 Vorteil erreicht, dass die Verarbeitung des Dokumentenbildes effizient durchgeführt werden kann. Die vorbestimmte Bildauflösung kann beispielsweise eine Bildauflösung von 640 x 480 Pixeln sein.

10 Gemäß einer Ausführungsform ist der Prozessor ausgebildet, das Dokumentenbild mittels eines Tiefpassfilters, insbesondere eines Gaußfilters, zu filtern. Dadurch wird der Vorteil erreicht, dass Bildrauschen in dem Dokumentenbild effizient reduziert werden kann. Durch die Filterung kann eine Weichzeichnung des Dokumentenbildes bewirkt werden.

15 Gemäß einer Ausführungsform ist der Prozessor ausgebildet, das Dokumentenbild mittels einer Segmentierung, insbesondere einer morphologischen Segmentierung oder mittels Toggle-Mapping, zu segmentieren, um die Mehrzahl von Bildsegmenten zu erhalten. Dadurch wird der Vorteil erreicht, dass homogen zusammenhängende Bereiche des Dokumentenbildes effizient bestimmt werden können.

20 Gemäß einer Ausführungsform ist der Prozessor ausgebildet, der Mehrzahl von Bildsegmenten, insbesondere mittels Labeling, jeweils eine Bildsegmentkennung zuzuordnen. Dadurch wird der Vorteil erreicht, dass die Mehrzahl von Bildsegmenten effizient adressiert werden kann.

25 Gemäß einer Ausführungsform ist der Prozessor ausgebildet, eine Mehrzahl von Bildsegmentmaßen der Mehrzahl von Bildsegmenten zu bestimmen, und die Mehrzahl von Bildsegmentmaßen mit einem vorbestimmten Textzeichenmaß zu vergleichen, um die Mehrzahl von Textzeichenbildsegmenten aus der Mehrzahl von Bildsegmenten auszuwählen. Dadurch wird der Vorteil erreicht, dass die Mehrzahl von  
30 Textzeichenbildsegmenten effizient ausgewählt werden können.

Das vorbestimmte Textzeichenmaß kann eine Fläche, ein Seitenverhältnis, einen Füllgrad oder eine Kompaktheit eines vorbestimmten repräsentativen Textzeichens anzeigen.

35 Gemäß einer Ausführungsform ist der Prozessor ausgebildet, einen Verbindungsgraph auf Basis der Mehrzahl von Textzeichenbildsegmenten zu erzeugen, und die Mehrzahl von Textzeichengruppen unter Verwendung des Verbindungsgraphen zu bestimmen. Dadurch wird der Vorteil erreicht, dass effiziente Ansätze der Graphentheorie zum Bestimmen der Mehrzahl von Textzeichengruppen eingesetzt werden können. Die

5    Schwerpunkte der Textzeichenbildsegmente können Knoten des Verbindungsgraphen bilden.

    Gemäß einer Ausführungsform ist der Prozessor ausgebildet, die Mehrzahl von Textzeichengruppen mittels einer Triangulation, insbesondere mittels einer Delaunay-  
10    Triangulation, des Verbindungsgraphen zu bestimmen. Dadurch wird der Vorteil erreicht, dass ein effizientes paarweises Gruppieren realisiert werden kann.

    Unter Verwendung der Triangulation, insbesondere der Delaunay-Triangulation, können Textzeichenbildsegmente mit gemeinsamen Eigenschaften bestimmt werden. Die  
15    bestimmten Textzeichenbildsegmente können eine Textzeichengruppe bilden, welche ein Wort oder eine Folge von Textzeichen repräsentiert. Die gemeinsamen Eigenschaften umfassen beispielsweise eine Höhe, einen Abstand, einen Lageversatz, eine Fläche, einen Winkel, einen Grauwert oder eine Strichstärke.

20    Gemäß einer Ausführungsform ist der Prozessor ausgebildet, eine Mehrzahl starker Zusammenhangskomponenten des Verbindungsgraphen zu bestimmen, und die Mehrzahl von Textzeichengruppen auf Basis der Mehrzahl starker Zusammenhangskomponenten zu bestimmen. Dadurch wird der Vorteil erreicht, dass ein effizientes paarweises Gruppieren realisiert werden kann. Zur Bestimmung der Mehrzahl  
25    starker Zusammenhangskomponenten kann beispielsweise der Algorithmus von Tarjan eingesetzt werden.

    Gemäß einer Ausführungsform ist der Prozessor ausgebildet, eine erste Textzeichengruppe und eine zweite Textzeichengruppe der Mehrzahl von  
30    Textzeichengruppen miteinander zu kombinieren, oder eine Textzeichengruppe der Mehrzahl von Textzeichengruppen in eine erste Textzeichengruppe und eine zweite Textzeichengruppe zu teilen. Dadurch wird der Vorteil erreicht, dass eine Kombination oder Teilung erreicht werden kann.

35    Die Kombination oder Teilung kann auf Basis von Eigenschaften der jeweiligen Textzeichengruppe durchgeführt werden. Dabei kann beispielsweise eine minimale Anzahl von Textzeichenbildsegmenten je Textzeichengruppe, ein maximaler oder minimaler Abstand, eine Ausrichtung, ein Grauwert, eine Fläche oder eine Strichstärke der Textzeichenbildsegmente berücksichtigt werden.

5 Gemäß einer Ausführungsform ist der Prozessor ausgebildet, die Mehrzahl von Anordnungsmaßen der Mehrzahl von Textzeichengruppen zu bestimmen. Dadurch wird der Vorteil erreicht, dass eine Mehrzahl von Anordnungsmaßen für einen Vergleich mit dem vorbestimmten Anordnungsmaß bereitgestellt werden kann.

10 Der Mehrzahl von Textzeichengruppen kann jeweils ein Anordnungsmaß zugeordnet sein. Ferner kann mehreren Textzeichengruppen, beispielsweise einem Paar von Textzeichengruppen, jeweils ein Anordnungsmaß zugeordnet sein.

15 Gemäß einer Ausführungsform ist der Prozessor ausgebildet, eine Raumlage der Bildkamera relativ zu dem erfassten Textbereich zu bestimmen, und den erfassten Textbereich auf Basis der bestimmten Raumlage perspektivisch zu entzerren. Dadurch wird der Vorteil erreicht, dass die Mehrzahl von Textzeichen des erfassten Textbereiches effizient erfasst werden kann.

20 Die Raumlage kann eine Pose der Bildkamera relativ zu dem erfassten Textbereich sein. Die Raumlage kann sechs Freiheitsgrade aufweisen, wobei drei Freiheitsgrade der relativen Position und drei Freiheitsgrade der relativen Ausrichtung zugeordnet sein können. Die perspektivische Entzerrung kann mittels einer Transformation durchgeführt werden, welche auf einer Schätzung einer Homographie basieren kann.

25 Gemäß einer Ausführungsform ist der Prozessor ausgebildet, eine optische Zeichenerkennung innerhalb des erfassten Textbereiches auf dem Identifikationsdokument durchzuführen. Dadurch wird der Vorteil erreicht, dass die Mehrzahl von Textzeichen des erfassten Textbereiches effizient erfasst werden kann.

30 Die optische Zeichenerkennung (Optical Character Recognition, OCR) kann einzeln für jedes Textzeichenbildsegment innerhalb des erfassten Textbereiches durchgeführt werden. Die optische Zeichenerkennung kann unter Verwendung eines hochaufgelösten ursprünglichen Dokumentenbildes durchgeführt werden.

35 Die optische Zeichenerkennung kann mittels einer Segmentierung, insbesondere mittels morphologischer Segmentierung oder mittels Toggle-Mapping, eines Textzeichenbildsegments durchgeführt werden. Das Textzeichenbildsegment kann in eine Mehrzahl von Kacheln unterteilt werden, wobei Eigenschaften der Mehrzahl von Kacheln  
40 in einem Merkmalsvektor zusammengefasst werden können. Durch einen Vergleich des

5 Merkmalsvektors mit einer Mehrzahl von vorbestimmten Vorlagenmerkmalsvektoren kann ein einzelnes Textzeichen erfasst werden.

Die morphologische Segmentierung ermöglicht dabei eine besonders effiziente Berechenbarkeit.

10

Gemäß einer Ausführungsform ist einem Textzeichen des Textbereiches ein vorbestimmter Zeichensatz zugeordnet, wobei der vorbestimmte Zeichensatz zulässige Buchstaben, Ziffern, oder Symbole des Textzeichens anzeigt, und wobei die optische Zeichenerkennung des Textzeichens unter Verwendung des vorbestimmten Zeichensatzes durchgeführt wird. Dadurch wird der Vorteil erreicht, dass eine effiziente optische Zeichenerkennung durchgeführt werden kann.

15

Der vorbestimmte Zeichensatz kann beispielsweise alphabetische, numerische oder alphanummerische Textzeichen umfassen. Für maschinen-lesbare Zonen kann der vorbestimmte Zeichensatz gemäß Standards der ICAO (International Civil Aviation Organization) standardisiert sein.

20

Gemäß einer Ausführungsform ist die Bildkamera ausgebildet, ein weiteres Bild des Identifikationsdokumentes zu erfassen, um ein weiteres Dokumentenbild zu erhalten, wobei der Prozessor ausgebildet ist, den Textbereich in dem weiteren Dokumentenbild zu erfassen, und die optische Zeichenerkennung innerhalb des erfassten Textbereiches ferner unter Verwendung des weiteren Dokumentenbildes durchzuführen. Dadurch wird der Vorteil erreicht, dass eine Fusion unter Verwendung des Dokumentenbildes und des weiteren Dokumentenbildes realisiert werden kann.

25

30

Gemäß einer Ausführungsform weist der Textbereich eine vorbestimmte Textbereichsanordnung gemäß eines vorbestimmten Textbereichsanordnungsmaßes auf dem Identifikationsdokument auf, wobei der Prozessor ausgebildet ist, das Identifikationsdokument in dem Dokumentenbild auf Basis des erfassten Textbereiches und des vorbestimmten Textbereichsanordnungsmaßes zu erfassen. Dadurch wird der Vorteil erreicht, dass das Identifikationsdokument in dem Dokumentenbild effizient erfasst werden kann.

35

Das vorbestimmte Textbereichsanordnungsmaß kann einen vorbestimmten Randabstand des Textbereiches bezüglich des Identifikationsdokumentes anzeigen. Auf Basis des

40

5 erfassten Textbereiches und des vorbestimmten Textbereichsanordnungsmaßes kann ein Suchbereich einer Kantendetektion des Identifikationsdokumentes in dem Dokumentenbild eingegrenzt werden.

10 Gemäß einer Ausführungsform ist der Prozessor ausgebildet, eine Raumlage der Bildkamera relativ zu dem erfassten Textbereich zu bestimmen, wobei die Bildkamera ausgebildet ist, ein weiteres Bild des Identifikationsdokumentes zu erfassen, um ein weiteres Dokumentenbild zu erhalten, und wobei der Prozessor ausgebildet ist, den Textbereich in dem weiteren Dokumentenbild unter Verwendung der bestimmten Raumlage zu erfassen. Dadurch wird der Vorteil erreicht, dass eine effiziente Erfassung und Verfolgung des Textbereiches realisiert werden kann.

Auf Basis der bestimmten Raumlage kann ein partielles Erfassen des Textbereiches und/oder ein partielles Erfassen der Textzeichen innerhalb des Textbereiches durchgeführt werden.

20

Gemäß einer Ausführungsform ist der Textbereich eine maschinen-lesbare Zone des Identifikationsdokumentes. Dadurch wird der Vorteil erreicht, dass ein standardisiertes vorbestimmtes Anordnungsmaß verwendet werden kann.

25 Die Anordnung der Mehrzahl von Textzeichen maschinen-lesbarer Zonen kann gemäß Standards der ICAO (International Civil Aviation Organization) standardisiert sein. Maschinen-lesbare Zonen können beispielsweise 3x30 Textzeichen, 2x36 Textzeichen, oder 2x44 Textzeichen aufweisen. Die Textzeichen können unter Verwendung einer OCR-B Schriftart gebildet sein.

30

Gemäß einer Ausführungsform kann die maschinen-lesbare Zone eine unregelmäßige Struktur aufweisen. Zur Auslesen bzw. zur Detektion von Daten einer derartigen maschinen-lesbaren Zone kann eine Analyse von beispielsweise Text-Gruppen durchgeführt werden, wobei die Berechnung der Koordinaten zur Entzerrung um die Bearbeitung bzw. Behandlung der jeweiligen Struktur erweitert werden kann.

35

Gemäß einem zweiten Aspekt betrifft die Erfindung ein Verfahren zum Erfassen eines Textbereiches auf einem Identifikationsdokument, wobei der Textbereich eine Mehrzahl von Textzeichen in einer vorbestimmten Anordnung gemäß einem vorbestimmten Anordnungsmaß aufweist, mit einem Erfassen eines Bildes des

40

5 Identifikationsdokumentes, um ein Dokumentenbild zu erhalten, einem Segmentieren des  
Dokumentenbildes, um eine Mehrzahl von Bildsegmenten zu erhalten, einem Auswählen  
einer Mehrzahl von Textzeichenbildsegmenten aus der Mehrzahl von Bildsegmenten,  
wobei die Textzeichenbildsegmente jeweils ein Textzeichen repräsentieren, einem  
Bestimmen einer Mehrzahl von Textzeichengruppen auf Basis der Mehrzahl von  
10 Textzeichenbildsegmenten, wobei die Textzeichengruppen jeweils eine Folge von  
Textzeichenbildsegmenten umfassen, und einem Vergleichen einer Mehrzahl von  
Anordnungsmaßen der Mehrzahl von Textzeichengruppen mit dem vorbestimmten  
Anordnungsmaß, um den Textbereich auf dem Identifikationsdokument zu erfassen.  
Dadurch wird der Vorteil erreicht, dass ein effizientes Konzept zum Erfassen eines  
15 Textbereiches auf einem Identifikationsdokument realisiert werden kann.

Das Verfahren kann durch das Mobilgerät ausgeführt werden. Weitere Merkmale des  
Verfahrens ergeben sich unmittelbar aus der Funktionalität des Mobilgerätes.

20 Gemäß einem dritten Aspekt betrifft die Erfindung ein Computerprogramm mit einem  
Programmcode zum Ausführen des Verfahrens, wenn das Computerprogramm auf einem  
Computer ausgeführt wird. Dadurch wird der Vorteil erreicht, dass das Verfahren  
automatisiert und wiederholbar ausgeführt werden kann. Das Mobilgerät kann  
programmtechnisch eingerichtet sein, um das Computerprogramm auszuführen.

25

Die Erfindung kann in Hardware und/oder Software realisiert werden.

Weitere Ausführungsbeispiele werden Bezug nehmend auf die beiliegenden Figuren  
näher erläutert. Es zeigen:

30

Fig. 1 ein Diagramm eines Mobilgerätes zum Erfassen eines Textbereiches auf einem  
Identifikationsdokument gemäß einer Ausführungsform;

Fig. 2 ein Diagramm eines Verfahrens zum Erfassen eines Textbereiches auf einem  
35 Identifikationsdokument gemäß einer Ausführungsform;

Fig. 3 ein Diagramm eines Textbereiches auf einem Identifikationsdokument gemäß  
einer Ausführungsform;

- 5 Fig. 4 ein Diagramm einer Mehrzahl von Textzeichen innerhalb von Textbereichen gemäß einer Ausführungsform;
- Fig. 5 ein Diagramm eines Verfahrens zum Erfassen eines Textbereiches auf einem Identifikationsdokument gemäß einer Ausführungsform;
- 10 Fig. 6 ein Diagramm eines Verfahrens zum Erfassen eines Textbereiches auf einem Identifikationsdokument gemäß einer Ausführungsform;
- Fig. 7 ein Diagramm perspektivisch verzerrter Textzeichen eines Textbereiches auf
- 15 einem Identifikationsdokument gemäß einer Ausführungsform;
- Fig. 8 ein Diagramm eines Verfahrens zum Erfassen eines Identifikationsdokumentes auf Basis eines erfassten Textbereiches gemäß einer Ausführungsform;
- 20 Fig. 9 ein Diagramm eines Verfahrens zum Verfolgen eines Textbereiches in einer Mehrzahl von Dokumentenbildern gemäß einer Ausführungsform; und
- Fig. 10 ein Diagramm eines Verfahrens zum Verfolgen eines Textbereiches in einer Mehrzahl von Dokumentenbildern gemäß einer Ausführungsform.
- 25 Fig. 1 zeigt ein Diagramm eines Mobilgerätes 100 zum Erfassen eines Textbereiches auf einem Identifikationsdokument gemäß einer Ausführungsform. Der Textbereich weist eine Mehrzahl von Textzeichen in einer vorbestimmten Anordnung gemäß einem vorbestimmten Anordnungsmaß auf.
- 30 Das Mobilgerät 100 umfasst eine Bildkamera 101, welche ausgebildet ist, ein Bild des Identifikationsdokumentes zu erfassen, um ein Dokumentenbild zu erhalten, und einen Prozessor 103, welcher ausgebildet ist, das Dokumentenbild zu segmentieren, um eine Mehrzahl von Bildsegmenten zu erhalten, eine Mehrzahl von Textzeichenbildsegmenten
- 35 aus der Mehrzahl von Bildsegmenten auszuwählen, wobei die Textzeichenbildsegmente jeweils ein Textzeichen repräsentieren, eine Mehrzahl von Textzeichengruppen auf Basis der Mehrzahl von Textzeichenbildsegmenten zu bestimmen, wobei die Textzeichengruppen jeweils eine Folge von Textzeichenbildsegmenten umfassen, und eine Mehrzahl von Anordnungsmaßen der Mehrzahl von Textzeichengruppen mit dem

5 vorbestimmten Anordnungsmaß zu vergleichen, um den Textbereich auf dem Identifikationsdokument zu erfassen.

Fig. 2 zeigt ein Diagramm eines Verfahrens 200 zum Erfassen eines Textbereiches auf einem Identifikationsdokument gemäß einer Ausführungsform. Der Textbereich weist eine  
10 Mehrzahl von Textzeichen in einer vorbestimmten Anordnung gemäß einem vorbestimmten Anordnungsmaß auf.

Das Verfahren 200 umfasst ein Erfassen 201 eines Bildes des Identifikationsdokumentes, um ein Dokumentenbild zu erhalten, ein Segmentieren 203 des Dokumentenbildes, um  
15 eine Mehrzahl von Bildsegmenten zu erhalten, ein Auswählen 205 einer Mehrzahl von Textzeichenbildsegmenten aus der Mehrzahl von Bildsegmenten, wobei die Textzeichenbildsegmente jeweils ein Textzeichen repräsentieren, ein Bestimmen 207 einer Mehrzahl von Textzeichengruppen auf Basis der Mehrzahl von Textzeichenbildsegmenten, wobei die Textzeichengruppen jeweils eine Folge von  
20 Textzeichenbildsegmenten umfassen, und ein Vergleichen 209 einer Mehrzahl von Anordnungsmaßen der Mehrzahl von Textzeichengruppen mit dem vorbestimmten Anordnungsmaß, um den Textbereich auf dem Identifikationsdokument zu erfassen. Das Verfahren 200 kann durch das Mobilgerät 100 ausgeführt werden.

25 Fig. 3 zeigt ein Diagramm eines Textbereiches 301 auf einem Identifikationsdokument 303 gemäß einer Ausführungsform. Ein Dokumentenbild 305 zeigt ein perspektivisch verzerrtes Bild des Identifikationsdokumentes 303.

Der Textbereich 301 kann eine maschinen-lesbare Zone des Identifikationsdokumentes  
30 303 sein. Die maschinen-lesbare Zone kann beispielsweise 3x30 Textzeichen, 2x36 Textzeichen, oder 2x44 Textzeichen aufweisen. Die Textzeichen können unter Verwendung einer OCR-B Schriftart gebildet sein.

Die maschinen-lesbare Zone kann jedoch auch eine unregelmäßige Struktur aufweisen.  
35 Zur Auslesen bzw. zur Detektion von Daten einer derartigen maschinen-lesbaren Zone kann eine Analyse von beispielsweise Text-Gruppen bzw. Bereichen durchgeführt werden.

Das Identifikationsdokument 303 kann eines der folgenden Identifikationsdokumente sein:  
40 Identitätsdokument, wie Personalausweis, Reisepass, Zugangskontrollausweis,

- 5 Berechtigungsausweis, Unternehmensausweis, Steuerzeichen oder Ticket, Geburtsurkunde, Führerschein, Kraftfahrzeugausweis, Zahlungsmittel, beispielsweise eine Bankkarte oder Kreditkarte. Das Identifikationsdokument 303 kann ferner einen elektronisch auslesbaren Schaltkreis, beispielsweise einen RFID-Chip umfassen. Das Identifikationsdokument 303 kann ein- oder mehrlagig bzw. papier- und/oder
- 10 kunststoffbasiert sein. Das Identifikationsdokument 303 kann aus kunststoffbasierten Folien aufgebaut sein, welche zu einem Kartenkörper mittels Verkleben und/oder Laminieren zusammengefügt werden, wobei die Folien bevorzugt ähnliche stoffliche Eigenschaften aufweisen.
- 15 Das Dokumentenbild 305 kann ein Farbwertbild oder ein Grauwertbild sein. Das Dokumentenbild 305 kann eine vorbestimmte Bildauflösung aufweisen.

Fig. 4 zeigt ein Diagramm einer Mehrzahl von Textzeichen innerhalb von Textbereichen 301 gemäß einer Ausführungsform. Die Textbereiche 301 sind jeweils durch eine

20 maschinen-lesbare Zone eines Identifikationsdokumentes gebildet. Maschinen-lesbare Zonen können unterschiedliche Strukturen aufweisen.

Es gibt drei verschiedene Typen, welche zwei oder drei Textzeilen umfassen. Die drei verschiedenen Typen können jeweils 90, 72 oder 88 einzelne Textzeichen aufweisen. Die

25 einzelnen Textzeichen können einem alphabetischen Zeichensatz, einem numerischen Zeichensatz oder einem alphanummerischen Zeichensatz entnommen sein.

Im Folgenden werden weitere Ausführungsbeispiele des Mobilgerätes 100 und des Verfahrens 200 näher beschrieben. Das Mobilgerät 100 und das Verfahren 200 können

30 zur Erfassung und Erkennung von Textbereichen 301, beispielsweise maschinen-lesbaren Zonen, in Echtzeit eingesetzt werden.

Eine Prüfung von Identifikationsdokumenten 303, wie beispielsweise von Reisepässen, ist eine bedeutende Aufgabe. Insbesondere in Situationen mit einem großen Durchsatz von

35 zu überprüfenden Personen, ist die Zeit zur Prüfung von Identifikationsdokumenten 303 sehr begrenzt. Textbereiche 301, insbesondere maschinen-lesbare Zonen, auf Identifikationsdokumenten 303 werden verwendet, um Prüfungen zu beschleunigen und um menschliche Fehler beim Lesen der Textzeichen, welche beispielsweise zur Identifikation verwendet werden, zu vermeiden.

5 Es gibt drei verschiedene Typen maschinen-lesbarer Zonen, welche üblicherweise auf der Vorderseite oder Identitätsseite von Identifikationsdokumenten 303 angeordnet sind. Die maschinen-lesbaren Zonen umfassen einen Absatz mit zwei oder drei parallelen Textzeilen schwarzen OCR-B Textes mit fester Breite und Größe sowie einem festen Zwischenzeilenabstand. Die Textzeilen umfassen persönliche Informationen über den  
10 Inhaber des Identifikationsdokumentes 303, Informationen über das Identifikationsdokument 303 und verschiedene Prüfsummen.

Das Erfassen und Auslesen von Textbereichen 301, beispielsweise maschinen-lesbarer Zonen, erfolgt üblicherweise unter Verwendung dedizierter stationärer Erfassungsgeräte  
15 oder Mobilgeräte. Bei üblichen Ansätzen wird der Textbereich 301, beispielsweise die maschinen-lesbare Zone, bezüglich des Gerätes manuell ausgerichtet bevor ein Erfassen und Auslesen durchgeführt werden kann. Dies verlängert die Erfassungs- und Auslesezeit und läuft der ursprünglichen Intention entgegen.

20 Dabei ist hervorzuheben, dass ein automatisches Erfassen und Auslesen eines Textbereiches 301, beispielsweise einer maschinen-lesbaren Zone, sowie eine optische Zeichenerkennung in einem allgemeinen Umfeld eine herausfordernde Aufgabe darstellen. Dadurch, dass kein Vorwissen über die Anwesenheit eines Textbereiches 301 vorhanden sein kann, sollte der Textbereich 301 automatisch und in Echtzeit identifiziert  
25 werden, ungeachtet einer möglichen Bewegungsunschärfe und anderer Widrigkeiten typischer Erfassungsszenarien mit Mobilgeräten. Die darauffolgende optische Zeichenerkennung soll dabei eine hohe Erkennungsleistung aufweisen.

Das Mobilgerät 100 und das Verfahren 200 stellen eine Lösung zur Erfassung und  
30 Erkennung von Textbereichen 301, beispielsweise maschinen-lesbarer Zonen, auf beliebigen Identifikationsdokumenten 303 in Echtzeit ohne zusätzliche Hardware bereit. Dabei kann eine hohe Varianz der Raumlagen oder Blickwinkel während der Erfassung toleriert werden. Ferner kann eine visuelle Rückmeldung über den Betriebsstatus an einen Benutzer bereitgestellt werden, um eine visuelle Führung zu ermöglichen. Im  
35 Folgenden wird eine mögliche Herangehensweise beschrieben.

Fig. 5 zeigt ein Diagramm eines Verfahrens 200 zum Erfassen eines Textbereiches auf einem Identifikationsdokument gemäß einer Ausführungsform. Das Verfahren 200 umfasst eine Mehrzahl von Schritten 501-513. Die Mehrzahl von Schritten 501-513 kann  
40 durch das Mobilgerät 100, beispielsweise den Prozessor 103, ausgeführt werden.

5

In Schritt 501 wird eine Textzeichen-Lokalisierung auf Basis eines Dokumentenbildes i durchgeführt. In Schritt 503 wird eine Textbereichs-Lokalisierung, beispielsweise eine MRZ-Lokalisierung, durchgeführt. In Schritt 505 wird eine Prüfung der Struktur des Textbereiches durchgeführt. In Schritt 507 wird eine Transformation geschätzt. In Schritt 10 509 erfolgen eine perspektivische Entzerrung und/oder ein Upsampling. In Schritt 511 wird eine optische Zeichenerkennung durchgeführt, welche Textzeichen-Vertrauenswerte bereitstellt. In Schritt 513 erfolgt eine Fusion, wobei der Textbereich, beispielsweise die MRZ, bereitgestellt werden kann. Die Struktur des Textbereiches, welche durch das vorbestimmte Anordnungsmaß angezeigt werden kann, wird innerhalb einer Mehrzahl von 15 Textzeichengruppen detektiert. Anschließend werden einzelne Textzeichen unter Verwendung einer geschätzten Transformation entzerrt und unter Verwendung einer optischen Zeichenerkennung verarbeitet. Mehrere Dokumentenbilder können fusioniert werden, um eine verbesserte Leistung zu erzielen.

20 Es wurde eine Menge von Texteigenschaften von Text auf Identifikationsdokumenten, insbesondere von maschinen-lesbaren Zonen, identifiziert, welche zur Erfassung und zum Auslesen hilfreich sind. Textbereiche auf Identifikationsdokumenten sind zumeist kleiner als textähnliche Verzerrungen im Hintergrund. Ein lokaler Textbereich, welcher Textzeichen umfasst, weist normalerweise eine einzige Farbe mit begrenzten Variationen 25 auf, und die Strichstärke jedes Textzeichens ist typischerweise konstant. Die Grenzen der Textzeichen sind üblicherweise geschlossen und Verbindungslinien auf der Kontur sind typischerweise glatt. Diese Grenzen können weitestgehend den Kanten entsprechen, welche in dem Dokumentenbild erfasst werden können. Einzelne Textzeichen innerhalb der Textbereiche haben zumeist sehr ähnliche Eigenschaften und sind entlang einer 30 gerichteten Line ausgerichtet. In den meisten Fällen kann eine minimale Anzahl von Textzeichen je Textbereich angenommen werden.

Der Ansatz zum mobilen Auslesen eines Textbereiches, beispielsweise einer maschinen-lesbaren Zone, umfasst mehrere Schritte 501-513. Zunächst wird die Lage von 35 Textzeichenbildsegmenten oder Textzeichenkandidaten in dem Dokumentenbild bestimmt. Aus dieser Information wird der Textbereich, beispielweise die maschinen-lesbare Zone, unter Berücksichtigung der räumlichen Anordnung zwischen Textzeichengruppen oder Textzeichengruppenkandidaten erfasst. Dann wird eine lokale Transformation für jedes Textzeichen geschätzt, welche zur Entzerrung verwendet 40 werden kann, gefolgt von einer Erkennung der Textzeichen, welche einen Vertrauenswert

5 bezüglich jedes Textzeichens der relevanten Teilmenge der OCR-B Schriftart bereitstellen kann. Schließlich wird Information aus mehreren Dokumentenbildern fusioniert, um die Ergebnisse zu verbessern. Im Folgenden werden diese Schritte detaillierter beschrieben.

Fig. 6 zeigt ein Diagramm eines Verfahrens 200 zum Erfassen eines Textbereiches auf  
10 einem Identifikationsdokument gemäß einer Ausführungsform. Das Verfahren 200 umfasst eine Mehrzahl von Schritten 601-611. Die Mehrzahl von Schritten 601-611 kann durch das Mobilgerät 100, beispielsweise den Prozessor 103, ausgeführt werden.

In Schritt 601 wird eine Vorverarbeitung eines Dokumentenbildes  $i$  durchgeführt. In Schritt  
15 603 erfolgt eine Segmentierung des Dokumentenbildes. In Schritt 605 wird ein Labeling durchgeführt. In Schritt 607 erfolgt eine Filterung. In Schritt 609 wird eine Gruppierung durchgeführt. In Schritt 611 erfolgt eine Teilung und/oder Kombination, um Textzeichengruppen bereitzustellen, welche Folgen oder Wörter repräsentieren. Zusammenhängende Textzeichenbildsegmente werden auf Basis einer initialen  
20 Segmentierung bestimmt, werden mittels eines Labeling verarbeitet, und werden gefiltert. Anschließend werden sie paarweise gruppiert und in Folgen oder Wörter aufgeteilt, und stellen somit eine Grundlage für die Erfassung des Textbereiches dar.

Die Erfassung des Textbereiches kann folgendermaßen erfolgen. Unter Verwendung  
25 eines Toggle-Mappings und eines zeit-linearen Labelings, kann eine initiale Erzeugung zusammenhängender Textzeichenbildsegmente erreicht werden. Ein initiales Filtern kann auf Grundlage der Geometrie und weiteren Randbedingungen von Textzeichen, wie beispielsweise Fläche, Ausdehnung, Seitenverhältnis, Füllgrad und Kompaktheit, durchgeführt werden. Ferner kann ein Kantenkontrast oder eine Strichstärke ausgewertet  
30 werden. Gemäß einer Ausführungsform wird dadurch ein Textzeichenmaß bereitgestellt.

Ähnliche Textzeichenbildsegmente können gruppiert werden, basierend auf  
Eigenschaften und einer räumlichen Kohärenz der Textzeichenbildsegmente. Aus  
Gründen der Effizienz kann eine Delaunay-Triangulation verwendet werden, um ein  
35 initiales paarweises Gruppieren zu erreichen, und somit Textzeichengruppen bereitzustellen. Paarweise Verbindungen in einem Verbindungsgraphen werden dann gefiltert unter Verwendung verschiedener relativer Kriterien, wie beispielsweise Höhe, Abstand, Lageversatz, Fläche, Winkel, Grauwert, und Strichstärke, gefolgt von einer Erzeugung starker zusammenhängender Komponenten. Dies ergibt eine Reihe  
40 geordneter Textzeichengruppen, welche im Idealfall einzelne Wörter oder Folgen von

5 Textzeichen repräsentieren, jedoch, abhängig von der Parametrierung und Struktur des Identifikationsdokumentes, mehrere Wörter oder Folgen von Textzeichen umfassen können. Deshalb kann eine zusätzliche Filterung eingesetzt werden.

Durch eine Teilung oder Kombination der Textzeichengruppen basierend auf  
10 Textzeichengruppeneigenschaften, wie beispielsweise minimaler Anzahl von Komponenten, maximaler/minimaler Abstand, Ausrichtung, Grauwert, Fläche und Strichstärke, können endgültige Textzeichengruppen bereitgestellt werden.

Aus den Textzeichengruppen können die individuellen Komponenten des Textbereiches,  
15 beispielsweise der maschinen-lesbaren Zone, durch eine Analyse ihrer Geometrie erfasst werden. Dabei kann nach Textzeichengruppen gesucht werden, die ein Kriterium minimaler Länge, beispielsweise 30 Textzeichen, erfüllen. Während der Auswahl können ferner ihre horizontalen und vertikalen Abstände analysiert werden. Gemäß einer Ausführungsform wird dadurch ein vorbestimmtes Anordnungsmaß bereitgestellt.

20

Folglich wird eine Anzahl von Textzeichengruppen bereitgestellt, welche den zu erfassenden Textbereich bilden. Die Anzahl von Textzeichengruppen kann in einer Verarbeitung durch eine optische Zeichenerkennung berücksichtigt werden. Gemäß einer Ausführungsform umfasst der zu erfassende Textbereich eine Textzeichengruppe.  
25 Gemäß einer weiteren Ausführungsform umfasst der zu erfassende Textbereich mehrere Textzeichengruppen.

Fig. 7 zeigt ein Diagramm perspektivisch verzerrter Textzeichen eines Textbereiches auf einem Identifikationsdokument gemäß einer Ausführungsform. Das Diagramm illustriert  
30 die Entzerrung von Textzeichen. Zunächst wird eine globale Transformation  $T$  unter Verwendung von Schnittpunkten angepasster Linien auf Basis von Bild-Koordinaten  $P_{c0I}$ ,  $P_{c1I}$ ,  $P_{c2I}$ ,  $P_{c3I}$  im Bild-Bereich und korrespondierenden Welt-Koordinaten  $P_{c0W}$ ,  $P_{c1W}$ ,  $P_{c2W}$ ,  $P_{c3W}$  im Welt-Bereich geschätzt. Anschließend kann eine lokale Transformation je Textzeichen geschätzt werden, welche dann für eine Entzerrung des  
35 Textzeichenbildsegmentes verwendet wird.

Die erfassten Textzeichen können unter Verwendung von Strukturinformationen des Textbereiches, beispielsweise der maschinen-lesbaren Zone, entzerrt werden. Zunächst werden horizontale und vertikale Linien auf die erfassten Textzeichenbildsegmente des  
40 Textbereiches unter Verwendung einer linearen Regression auf ihre Schwerpunkte

5 angepasst. Diese Linien werden ferner geschnitten, um verbesserte Schätzungen der vier  
äußersten Textzeichenmittelpunkte  $P_{ci}$  bereitzustellen. Unter Verwendung bekannter  
Eigenschaften der Schriftart, beispielsweise der OCR-B Schriftart, können  
korrespondierende Koordinaten  $P_{cw}$  in dem entzerrten Welt-Bereich bestimmt werden,  
welche es erlauben, eine perspektivische Transformation  $T$  zu schätzen.

10

Für jeden Textzeichenschwerpunkt, wie aus dem Schnitt-Prozess erhalten, können die  
Grenzen des Textzeichenbildsegmentes im Welt-Bereich unter Verwendung von  
Schrifteigenschaften bestimmt werden und dann in das Dokumentenbild projiziert werden.  
Nun kann eine lokale Transformation für jedes Textzeichen geschätzt werden, welche für  
15 die Entzerrung eingesetzt werden kann. Um die Eingangsdaten für die optische  
Zeichenerkennung zu verbessern, kann ferner ein Upsampling der  
Textzeichenbildsegmente während der Entzerrung durchgeführt werden.

Die optische Zeichenerkennung kann folgendermaßen durchgeführt werden. Die optische  
20 Zeichenerkennung verwendet die Ergebnisse einer folgenden Binarisierung als  
Eingangsdaten. Hierfür kann beispielsweise Toggle-Mapping verwendet werden, wobei  
die erhaltene Binärmaske einem Labeling unterzogen werden kann, und eine minimale  
Begrenzungsbox für das Textzeichen geschätzt werden kann.

25 Durch eine Auswahl von Dokumentenbildern kann eine kleine Anzahl von Abtastwerten  
für die Erkennung eines einzelnen Textzeichens ausreichen. Dabei kann eine Überlapp-  
Metrik zur Zeichenerkennung verwendet werden, welche auf einem regulären Gitter  
berechnet werden kann. Der lokale Überlapp kann für jede Kachel oder Zelle des  
Textzeichenbildsegmentes berechnet werden und als Merkmalsvektor gespeichert  
30 werden. Unter Verwendung eines  $L_1$  Abstandes, kann eine Ähnlichkeit hinsichtlich einer  
Mehrzahl von Vorlagen berechnet werden, welche auch als Vertrauenswert aufgefasst  
werden kann. Initiale Abtastwerte können durch ein Rendern von True-Type Schriften  
erzeugt werden. Anschließend kann eine kleine Anzahl von realen Abtastwerten  
hinzugefügt werden, welche unter Verwendung des beschriebenen Ansatzes extrahiert  
35 wurde.

Eine Fusion kann beispielsweise folgendermaßen durchgeführt werden. Wenn ein Live-  
Video zur Verfügung steht, können mehrere Dokumentenbilder  $i$  durch das Mobilgerät  
verarbeitet werden, um die Robustheit zu erhöhen. Für einen folgenden Fusions-Prozess  
40 sollten Korrespondenzen zwischen den Textzeichen hergestellt werden. Beim Verfolgen

- 5 durch Erfassen wird die Struktur des initialen Erfassungsergebnisses berücksichtigt, wann immer nach geeigneten Dokumentbildern gesucht wird. Gemäß einer Ausführungsform werden Dokumentenbilder zur Fusion auf Basis einer Bildschärfe oder Beleuchtung der Dokumentenbilder ausgewählt.
- 10 In jedem Dokumentenbild  $i$ , für jedes Textzeichen  $j$ , beispielsweise der maschinenlesbaren Zone, können Abstände  $d_{i,j,k}$  betreffend aller bekannten Vorlagen  $k$  erfasst werden. Für jeden Eintrag kann der Mittelwert hinsichtlich aller Dokumentenbilder berechnet werden gemäß  $d_{j,k} = \text{mean}(d_{i,j,k})$ . Das Endergebnis je Textzeichen kann dann berechnet werden als jedes, welches den kleinsten Abstand  $q_i = \max(q_{i,k})$  aufweist.
- 15 Gemäß einer Ausführungsform ist  $q = 1-d$ .

Der beschriebene Ansatz zum Erfassen und Auslesen eines Textbereiches auf einem Identifikationsdokument, beispielsweise einer maschinenlesbaren Zone, in Echtzeit operiert ohne eine genaue Ausrichtung des Identifikationsdokumentes oder des

20 Textbereiches. Durch eine initiale Erfassung des Textbereiches und eine Fusion von Ergebnissen mehrerer Dokumentenbilder, kann die optische Zeichenerkennung hochwertige Ergebnisse bereitstellen. Da auf eine genaue Ausrichtung verzichtet werden kann, wird Zeit eingespart. Ferner kann der Textbereich schnell extrahiert werden.

25 Der Ansatz kann auf verschiedene Arten verbessert werden. Wenn zusätzliche Textzeichentrainingsdaten zugänglich sind, kann der Abgleich mit Vorlagen durch einen geeigneten Klassifizierer ersetzt werden. Dies kann die Erfassungsergebnisse oder Ausleseergebnisse des Textbereiches sowie die Laufzeit verbessern. Ein Verfolgen des Textbereiches kann die Robustheit erhöhen, da weitere Eingangsdaten für die optische

30 Zeichenerkennung zur Verfügung stehen. Ferner kann eine leichte Biegung des Identifikationsdokumentes berücksichtigt werden.

Gemäß einer Ausführungsform kann die Kombination von Einzelbildern unmittelbar in der Bildebene oder der Regionsebene erfolgen. In diesem Fall könnte der OCR-Schritt pro

35 Zeichen ein einziges Mal durchgeführt werden.

Im Folgenden wird ein Ansatz zur Erfassung eines Identifikationsdokumentes mittels eines vorab erfassten Textbereiches beschrieben. Anschließend wird das Verfolgen (engl. Tracking) von Textbereichen erläutert. Das Verfolgen erlaubt ein kontinuierliches Erfassen

5 oder Auslesen und kann zur Überwachung der Erfassungssituation dienen. Schließlich wird auf das partielle Erfassen oder Auslesen von Textbereichen eingegangen.

Fig. 8 zeigt ein Diagramm eines Verfahrens zum Erfassen eines Identifikationsdokumentes auf Basis eines erfassten Textbereiches gemäß einer  
10 Ausführungsform. Das Verfahren umfasst eine Mehrzahl von Schritten 801-813. Die Mehrzahl von Schritten 801-813 kann durch das Mobilgerät 100, beispielsweise den Prozessor 103, ausgeführt werden. Das Verfahren 200 zum Erfassen eines Textbereiches auf einem Identifikationsdokument kann das Verfahren zum Erfassen eines Identifikationsdokumentes auf Basis eines erfassten Textbereiches umfassen.

15

In Schritt 801 wird eine Transformation auf Basis von Textbereichslinien geschätzt. In Schritt 803 erfolgt eine Textbereichserweiterung. In Schritt 805 wird eine Kantendetektion durchgeführt. In Schritt 807 erfolgt eine Liniensegmenterfassung. In Schritt 809 wird eine H/V Profilberechnung durchgeführt. In Schritt 811 erfolgt eine Spitzen- oder Peak-  
20 Erfassung. In Schritt 813 wird eine Fusion durchgeführt, um Informationen bezüglich der Ecken des Identifikationsdokumentes bereitzustellen.

Die Erfassung des Identifikationsdokumentes, beispielsweise der Grenzen des Identifikationsdokumentes, kann folgendermaßen durchgeführt werden. Es kann ein  
25 allgemeiner Ansatz zur Erfassung eines perspektivisch verzerrten Rechtecks eingesetzt werden. Dabei können nach anfänglicher oder initialer Filterung von hochfrequenten Störungen zunächst Linien erfasst und danach paarweise zu Hypothesen zusammengefasst werden.

30 Danach kann eine Bewertung der Hypothesen je nach Beschaffenheit eines Kantenbildes entlang ausgewählter Grenzen des Identifikationsdokumentes erfolgen. Bei diesem Ansatz sollte sich das Identifikationsdokument innerhalb eines vordefinierten und durch den Benutzer einstellbaren Bereiches befinden, was einer groben Ausrichtung von Identifikationsdokument und Mobilgerät entsprechen kann.

35

Wenn ein Textbereich, beispielsweise eine maschinen-lesbare Zone, auf dem Identifikationsdokument vorhanden ist und deren Begrenzungen bekannt sind, kann dies als Vorwissen zur Erfassung des Identifikationsdokumentes, beispielsweise der Grenzen des Identifikationsdokumentes, verwendet werden. Zusammen mit einer Anwendung des  
40 beschriebenen Ansatzes auf eine nur gering eingeschränkte Raumlage bei der Erfassung,

5 kann das Identifikationsdokument weitgehend automatisch während der Erfassung des Textbereiches erfasst werden.

In jedem einzelnen Dokumentenbild  $F_i$  wird zunächst der Textbereich erfasst und eine Transformation  $T$  geschätzt. Dabei werden die Schwerpunkte  $C_{llc0}$  der zum Textbereich  
10 gehörigen Textzeichengruppen im Bild-Bereich jeweils zeilenweise und spaltenweise durch eine Gerade approximiert. Es ergeben sich beispielsweise  $L$  horizontale und  $C$  vertikale Geraden je Textbereich, aus welchen neue Schwerpunkte  $C_{llc1}$  geschätzt werden können. Da die Anordnung oder Geometrie der Textzeichen innerhalb des Textbereiches je nach Typ vorbekannt sein kann, beispielsweise  $2 \times 44$ ,  $2 \times 36$  oder  $2 \times 30$  Textzeichen, sind  
15 Koordinaten  $C_{wlc}$  für das Modell gegeben und aus den Korrespondenzen  $\{C_{llc1}, C_{wlc}\}$  kann schließlich die Transformation  $T$  geschätzt werden.

Falls Fluchtpunkte verwendet werden, kann eine einheitliche horizontale und/oder vertikale Teilung geschätzt werden. Die erhaltene Transformation kann unmittelbar zur  
20 Entzerrung des Dokumentenbildes  $F_i$  verwendet werden, wobei eine grobe Ausdehnung ausgehend von dem Textbereich, beispielsweise links, rechts, oben oder unten, wiederum durch eine Standardisierung oder Spezifikation des Identifikationsdokumentes gegeben sein kann. In diesem entzerrten und auf eine vorbestimmte Ausdehnung skalierten Dokumentenbild  $F_{ir}$  kann nun nach Grenzen des Identifikationsdokumentes gesucht  
25 werden.

In den relevanten Bereichen des Dokumentenbildes  $F_{imap}$  kann nun eine Kantendetektion durchgeführt werden. Dann werden horizontale und vertikale Liniensegmente, beispielsweise mit einer minimalen Länge von 3 Pixeln, erfasst. Auf den  
30 Zwischendokumentenbildern  $F_{imapv}$  und  $F_{imaph}$  können horizontale  $h_{imap}$  und vertikale  $v_{imap}$  Projektionsprofile berechnet werden. Mit dieser Information kann nun nach Spitzen, Peaks oder Maxima im durch die Spezifikation gegebenen Bereich gesucht werden. Die Auswahl kann unter Verwendung eines geeigneten Schwellwertes erfolgen. Zur Verbesserung der Robustheit können bei der Analyse der Projektionsprofile die Ergebnisse von mehreren  
35 einzelnen Dokumentenbildern  $F_i$  verwendet werden. Dabei erfolgt eine Mittelung der beteiligten Profile beispielsweise gemäß  $h_{imap} = 1/n(h_{imap})$ ;  $v_{imap} = 1/n(v_{imap})$ ;  $i=1 \dots n$ .

Aus den vorherigen Teilschritten ergeben sich jeweils zwei Indizes für die Profile  $h_{imap}$ ,  $v_{imap}$ . Durch Anwendung auf das Dokumentenbild  $F_{ir}$  ergibt sich ein entzerrtes  
40 Dokumentenbild, welches für weitere Analyseschritte oder auch unmittelbar zur

5 Verfolgung verwendet werden kann. Da das Dokumentenbild  $F_{ir}$  durch eine bekannte Transformation ermittelt wurde, sind damit auch die Grenzen des Identifikationsdokumentes in den ursprünglichen Dokumentenbildern bekannt. Im Falle der Fusionierung wird aus jenem Dokumentenbild extrahiert, das bezüglich der ermittelten Indizes die Summe der größten Einträge in den korrespondierenden Profilen aufweist.

10

Fig. 9 zeigt ein Diagramm eines Verfahrens zum Verfolgen eines Textbereiches in einer Mehrzahl von Dokumentenbildern gemäß einer Ausführungsform. Das Verfahren umfasst eine Mehrzahl von Schritten 901-913. Die Mehrzahl von Schritten 901-913 kann durch das Mobilgerät 100, beispielsweise den Prozessor 103, ausgeführt werden. Das Verfahren 200 zum Erfassen eines Textbereiches auf einem Identifikationsdokument kann das Verfahren zum Verfolgen eines Textbereiches in einer Mehrzahl von Dokumentenbildern umfassen.

Die Schritte 901-905 zeigen den grundsätzlichen Ablauf bei der Verfolgung des Textbereiches. In Schritt 901 wird eine Erfassung des Textbereichs durchgeführt, wobei eine Raumlage bestimmt wird. In Schritt 903 erfolgt eine Verfolgung des Textbereichs auf Basis der bestimmten Raumlage. In Schritt 905 wird das Ergebnis der Verfolgung des Textbereichs geprüft, wobei bei erfolgreicher Verfolgung des Textbereichs mit Schritt 903 fortgefahren wird, und wobei bei nicht erfolgreicher Verfolgung des Textbereichs mit Schritt 901 fortgefahren wird.

Die Schritte 907-913 können innerhalb der Verfolgung des Textbereichs, beispielsweise in Schritt 903, ausgeführt werden. In Schritt 907 erfolgt eine Initialisierung auf Basis eines Dokumentenbildes und eines Erfassungsergebnisses. In Schritt 909 wird eine Verfolgung auf Basis des Dokumentenbildes durchgeführt, wobei eine Raumlage bestimmt werden kann. In Schritt 911 erfolgt eine Re-Initialisierung auf Basis des Erfassungsergebnisses. In Schritt 913 wird eine OCR Entzerrung auf Basis des Dokumentenbildes durchgeführt, um Textzeichenbildsegmente oder Textzeichen-Bildbereiche bereitzustellen. Die Initialisierung und Re-Initialisierung in den Schritten 907, 911 erfolgen auf Basis des Erfassungsergebnisses. Gemäß einer Ausführungsform wird bei der Verfolgung nur das aktuelle Dokumentenbild verarbeitet. In Schritt 913 können OCR-Daten aus dem aktuellen Dokumentenbild erzeugt werden.

Fig. 10 zeigt ein Diagramm eines Verfahrens zum Verfolgen eines Textbereiches in einer Mehrzahl von Dokumentenbildern gemäß einer Ausführungsform. Das Verfahren umfasst

5 eine Mehrzahl von Schritten 1001-1017. Die Mehrzahl von Schritten 1001-1017 kann durch das Mobilgerät 100, beispielsweise den Prozessor 103, ausgeführt werden. Das Verfahren 200 zum Erfassen eines Textbereiches auf einem Identifikationsdokument kann das Verfahren zum Verfolgen eines Textbereiches in einer Mehrzahl von Dokumentenbildern umfassen.

10

Die Schritte 1001-1017 verdeutlichen die Vorgänge bei der Verfolgung eines einzelnen Dokumentenbildes. In Schritt 1001 wird ein Bewegungsmodell angewendet. In Schritt 1003 wird eine Sichtbarkeitsprüfung durchgeführt. In Schritt 1005 erfolgt eine Textzeichenauswahl. In Schritt 1007 wird eine Entzerrung durchgeführt. In Schritt 1009 erfolgt eine Segmentierung. In Schritt 1011 wird ein Labeling durchgeführt. In Schritt 1013 erfolgt ein Abgleich. In Schritt 1015 wird eine Ausreißer-Entfernung durchgeführt. In Schritt 1017 erfolgt eine Schätzung der Raumlage.

Um eine schnelle Rückmeldung für einen Benutzer zu erlauben, kann die Erfassung des Textbereiches mit einer relativ geringen Bildauflösung arbeiten. Zudem sollte der Textbereich zur Verarbeitung vollständig sichtbar sein, da möglicherweise keine Vorkenntnis von Dokumentenbild zu Dokumentenbild transferiert wird. Eine Verfolgung des Textbereiches von Dokumentenbild zu Dokumentenbild kann jedoch ein kontinuierliches Erfassen oder Auslesen von Textzeichen erlauben. Zudem kann die Erfassungssituation besser überwacht werden, da Bewegungen des Benutzers unmittelbar ausgewertet werden können und dadurch eine Rückmeldung und/oder Anleitung durch das Mobilgerät erfolgen kann.

Textbereiche können mit Ansätzen zur Verfolgung natürlicher Merkmale (engl. Natural Features) oftmals nicht hinreichend verfolgt werden. Deshalb wird ein Ansatz zur Verfolgung des Textbereiches mit sechs Freiheitsgraden, basierend auf einzelnen Textzeichen, eingesetzt. Zunächst erfolgt eine einmalige Erfassung des Textbereiches, um den Verfolger zu initialisieren. Mit der bekannten Transformation zwischen Bild-Bereich und Welt-Bereich werden die einzelnen Textzeichenbildsegmente entzerrt, segmentiert und abgelegt. Außerdem kann eine erste Raumlage oder Pose berechnet werden.

Zur Laufzeit kann zunächst die vorhergehende Raumlage verwendet werden, um zusammen mit intrinsischen Parametern der Bildkamera eine Welt-Position eines Referenzbereiches in das Dokumentenbild zu projizieren. Unter Beachtung der Grenzen

5 des Dokumentenbildes lassen sich somit Rückschlüsse auf die Sichtbarkeit treffen. Davon können pro Textzeichengruppe oder Linie des Textbereiches  $n \geq 2$  sichtbare Referenzbereiche ausgewählt werden, deren neue Position im Dokumentenbild gesucht wird. Dabei kann lokal eine neue Transformation  $T_{lc}$  berechnet werden, womit der korrespondierende Bereich im Dokumentenbild entzerrt werden kann. Nach einer  
10 Segmentierung werden die  $m$  größten zusammenhängenden Bereiche ausgewählt. Der jeweilige Bereich wird dann auf eine einheitliche Ausdehnung gebracht und mittels eines Überlapp-Deskriptors, analog zur Zeichenerkennung, in einen Merkmalsvektor konvertiert, welcher mit den Daten der aktuellen Referenzregion  $R_{lc}$  verglichen werden kann. Bei einer Übereinstimmung wird das zugehörige Zentrum des Bereiches als neue Position  
15 abgelegt. Mit allen  $n \geq 2$  Korrespondenzen je Linie, mit mindestens 2 Linien, kann mittels eines RANSAC-Ansatzes eine Homographie geschätzt werden. Bei Erfolg wird die Schätzung der Raumlage aus den gefilterten Erfassungen, sogenannte Inliers, unter Verwendung aktueller intrinsischer Parameter der Bildkamera durchgeführt.

20 Zur Verbesserung von Robustheit und Laufzeit kann ein Bewegungsmodell verwendet werden. Dabei wird basierend auf den letzten Erfassungen  $P_i, P_{i-1}$  die aktuelle Raumlage entsprechend geschätzt, beispielsweise durch eine Mittelung der Änderungen von Translation und/oder Rotation.

25 Falls die Verfolgung abbricht, kann mittels einer erneuten Erfassung des Textbereiches die Raumlage erneut bestimmt werden, ohne dass die Referenzdaten aktualisiert werden. Die Extraktion der entzerrten Textzeichenbildsegmente oder Patch-Daten erfolgt unter Verwendung der aktuellen Raumlage und unter Berücksichtigung der Sichtbarkeit analog zur Erfassung des Textbereiches und kann auf die beschriebene Weise durchgeführt  
30 werden. Der Textbereich könnte selbst bei der Initialisierung nicht vollständig sichtbar sein. Dann kann eine bekannte Raumlage verwendet werden. Fehlende Referenzdaten können während der Verfolgung hinzugefügt werden.

Ein partielles Erfassen oder Auslesen des Textbereiches kann beispielsweise  
35 folgendermaßen realisiert werden. Wenn die Textbereichsanordnung oder Lage eines vollständig oder unvollständig sichtbaren Textbereiches auf dem Identifikationsdokument bekannt ist, können einzelne Textzeichen gemäß ihrer Sichtbarkeit und optional gemäß einer Konfidenz bezüglich von Referenzdaten erfasst oder ausgelesen werden. Damit sind eine lokale Rückmeldung und eine schnellere Erfassung möglich, falls einzelne  
40 Textzeichen der Konfidenzprüfung zunächst nicht genügen.

5

Während typischerweise der gesamte Textbereich erfasst oder ausgelesen wird, kann eine Modifikation zum Erfassen oder Auslesen von Teilresultaten in Teiltextbereichen durchgeführt werden. Es kann je Textzeichen eine Liste von Erfassungen oder Auslesungen verwaltet werden, welche eine Konfidenz bezüglich bekannter Referenzdaten anzeigen kann. Damit kann entschieden werden, ob eine ausreichende Konfidenz oder Abdeckung der räumlichen Lage mit optionaler Orientierungskarte (engl. Orientation Map) vorliegt. Dies kann entsprechend visualisiert werden, sodass der Benutzer angeleitet werden kann, Textzeichen mit geringer Konfidenz nochmals zu erfassen. Damit können auch mögliche lokale Modifikationen des Textbereiches zum Zweck der Prüfung des Identifikationsdokumentes auf Basis der Konfidenz visualisiert werden. Durch Verwendung einer Orientierungskarte kann ein nicht erwünschter Einfluss der Erfassungssituation auf die finale Konfidenz eingeschränkt werden.

Folglich können Erweiterungen für die mobile Auswertung von Identifikationsdokumenten mit Textbereichen, beispielsweise maschinen-lesbaren Zonen, eingesetzt werden, welche es ermöglichen, die Grenzen des Identifikationsdokumentes zu bestimmen, den Textbereich von Dokumentenbild zu Dokumentenbild zu verfolgen, sowie den Inhalt des Textbereiches partiell zu erfassen oder auszulesen. Das Mobilgerät kann ein Smartphone sein. Im Falle der Berechnung von Teilen des Verfolgungsansatzes mit mobilen Grafikprozessoren, ist ein weiterer Leistungsgewinn zu erwarten.

Es wird ein effizienter Ansatz zum Erfassen oder Auslesen von Textbereichen mit einem Mobilgerät auf Basis von Dokumentenbildern eingesetzt. Dabei wird das Identifikationsdokument nicht manuell ausgerichtet und es werden keine weiteren Annahmen über die Anordnung des Textbereiches auf dem Identifikationsdokument eingebracht. Optional kann der Textbereich verfolgt werden, sodass je nach Sichtbarkeit und Konfidenz partiell erfasst oder ausgelesen werden kann. Nach Erfassung einiger Dokumentenbilder kann ein Fusionsschritt bezüglich der optischen Zeichenerkennung erfolgen zum Erfassen oder Auslesen des Textbereiches und optional zur Bestimmung der Grenzen des Identifikationsdokumentes.

Die optische Zeichenerkennung kann auf einer Binarisierung mit einem folgenden Abtrennen eines Textzeichenbereichs basieren. Unter Verwendung von Trainingsdaten kann das Abtrennen verbessert werden. Dabei kann ein Ansatz maschinellen Lernens, beispielsweise eine Stützvektormaschine (Support Vector Machine, SVM), unter

5 Verwendung voller Informationen der Textzeichenbildsegmente eingesetzt werden. Die Fusion unter Verwendung von Dokumentenbildern kann dahingehend modifiziert werden, Abtastwerte aus verschiedenen Blickrichtungen zu verwenden. Dabei kann eine leichte Zitterbewegung, welche in mobilen Erfassungsszenarien vorhanden sein kann, für diesen Zweck ausreichen.

10

Es werden mehrere Vorteile erreicht. Zu einen kann auf eine genaue manuelle Ausrichtung, beispielsweise eine manuelle orthogonale Ausrichtung, des Textbereiches relativ zu der Bildkamera verzichtet werden. Zudem kann einem Benutzer eine visuelle Rückmeldung über den Status der Erfassung des Textbereiches in Echtzeit bereitgestellt werden. Die Erfassungsergebnisse können automatisch, sowohl vollständig als auch

15 partiell, fusioniert werden, um die Robustheit zu erhöhen.

Es besteht ferner die Möglichkeit, die Grenzen des Identifikationsdokumentes auf Basis der Erfassung des Textbereiches zu lokalisieren. Dabei können die Grenzen des Identifikationsdokumentes auf Basis mehrerer Erfassungsergebnisse des Textbereiches

20 bestimmt werden. Die Erfassungsbedingungen können mittels eines Verfolgens des Textbereiches überwacht werden. Beispielsweise kann eine Überwachung eines Abstandes oder einer perspektivischen Verzerrung erfolgen. Ferner kann ein Verfolgen und ein partielles Erfassen oder Auslesen des Textbereiches durchgeführt werden.

25

Folglich wird ein mobiles Erfassen oder Auslesen von Textbereichen mittels Mobilgeräten ohne zusätzliche Hardware und ohne eine manuelle Ausrichtung des Identifikationsdokumentes bezüglich der Bildkamera erreicht. Dabei kann das Identifikationsdokument unter Verwendung der Erfassungsergebnisse des Textbereiches

30 als Vorabinformation in dem Dokumentenbild erfasst werden.

**5 BEZUGSZEICHENLISTE**

	100	Mobilgerät
	101	Bildkamera
	103	Prozessor
10		
	200	Verfahren
	201	Erfassen
	203	Segmentieren
	205	Auswählen
15	207	Bestimmen
	209	Vergleichen
	301	Textbereich
	303	Identifikationsdokument
20	305	Dokumentenbild
	501-513	Schritte
	601-611	Schritte
	801-813	Schritte
25	901-913	Schritte
	1001-1017	Schritte

## 5 PATENTANSPRÜCHE

1. Mobilgerät (100) zum Erfassen eines Textbereiches (301) auf einem Identifikationsdokument (303), wobei der Textbereich (301) eine Mehrzahl von Textzeichen in einer vorbestimmten Anordnung gemäß einem vorbestimmten Anordnungsmaß aufweist, mit:

einer Bildkamera (101), welche ausgebildet ist, ein Bild des Identifikationsdokumentes (303) zu erfassen, um ein Dokumentenbild (305) zu erhalten; und

15 einem Prozessor (103), welcher ausgebildet ist:

- das Dokumentenbild (305) zu segmentieren, um eine Mehrzahl von Bildsegmenten zu erhalten;
- eine Mehrzahl von Textzeichenbildsegmenten aus der Mehrzahl von Bildsegmenten auszuwählen, wobei die Textzeichenbildsegmente jeweils ein Textzeichen repräsentieren;
- eine Mehrzahl von Textzeichengruppen auf Basis der Mehrzahl von Textzeichenbildsegmenten zu bestimmen, wobei die Textzeichengruppen jeweils eine Folge von Textzeichenbildsegmenten umfassen; und
- eine Mehrzahl von Anordnungsmaßen der Mehrzahl von Textzeichengruppen mit dem vorbestimmten Anordnungsmaß zu vergleichen, um den Textbereich (301) auf dem Identifikationsdokument (303) zu erfassen.

2. Mobilgerät (100) nach Anspruch 1, wobei der Prozessor (103) ausgebildet ist, das Dokumentenbild (305) mittels einer Segmentierung, insbesondere einer morphologischen Segmentierung oder mittels Toggle-Mapping, zu segmentieren, um die Mehrzahl von Bildsegmenten zu erhalten.

3. Mobilgerät (100) nach einem der vorstehenden Ansprüche, wobei der Prozessor (103) ausgebildet ist, eine Mehrzahl von Bildsegmentmaßen der Mehrzahl von Bildsegmenten zu bestimmen, und die Mehrzahl von Bildsegmentmaßen mit einem vorbestimmten Textzeichenmaß zu vergleichen, um die Mehrzahl von Textzeichenbildsegmenten aus der Mehrzahl von Bildsegmenten auszuwählen.

4. Mobilgerät (100) nach einem der vorstehenden Ansprüche, wobei der Prozessor (103) ausgebildet ist, einen Verbindungsgraph auf Basis der Mehrzahl von

- 5 Textzeichenbildsegmenten zu erzeugen, und die Mehrzahl von Textzeichengruppen unter Verwendung des Verbindungsgraphen zu bestimmen.
5. Mobilgerät (100) nach Anspruch 4, wobei der Prozessor (103) ausgebildet ist, die Mehrzahl von Textzeichengruppen mittels einer Triangulation, insbesondere mittels einer  
10 Delaunay-Triangulation, des Verbindungsgraphen zu bestimmen.
6. Mobilgerät (100) nach einem der vorstehenden Ansprüche, wobei der Prozessor (103) ausgebildet ist, eine erste Textzeichengruppe und eine zweite Textzeichengruppe der Mehrzahl von Textzeichengruppen miteinander zu kombinieren, oder eine  
15 Textzeichengruppe der Mehrzahl von Textzeichengruppen in eine erste Textzeichengruppe und eine zweite Textzeichengruppe zu teilen.
7. Mobilgerät (100) nach einem der vorstehenden Ansprüche, wobei der Prozessor (103) ausgebildet ist, eine Raumlage der Bildkamera (101) relativ zu dem erfassten  
20 Textbereich (301) zu bestimmen, und den erfassten Textbereich (301) auf Basis der bestimmten Raumlage perspektivisch zu entzerren.
8. Mobilgerät (100) nach einem der vorstehenden Ansprüche, wobei der Prozessor (103) ausgebildet ist, eine optische Zeichenerkennung innerhalb des erfassten  
25 Textbereiches (301) auf dem Identifikationsdokument (303) durchzuführen.
9. Mobilgerät (100) nach Anspruch 8, wobei einem Textzeichen des Textbereiches (301) ein vorbestimmter Zeichensatz zugeordnet ist, wobei der vorbestimmte Zeichensatz zulässige Buchstaben, Ziffern, oder Symbole des Textzeichens anzeigt, und wobei die  
30 optische Zeichenerkennung des Textzeichens unter Verwendung des vorbestimmten Zeichensatzes durchgeführt wird.
10. Mobilgerät (100) nach Anspruch 8 oder 9, wobei die Bildkamera (101) ausgebildet ist, ein weiteres Bild des Identifikationsdokumentes (303) zu erfassen, um ein weiteres  
35 Dokumentenbild zu erhalten, und wobei der Prozessor (103) ausgebildet ist, den Textbereich (301) in dem weiteren Dokumentenbild zu erfassen, und die optische Zeichenerkennung innerhalb des erfassten Textbereiches (301) ferner unter Verwendung des weiteren Dokumentenbildes durchzuführen.

5 11. Mobilgerät (100) nach einem der vorstehenden Ansprüche, wobei der Textbereich  
(301) eine vorbestimmte Textbereichsanordnung gemäß eines vorbestimmten  
Textbereichsanordnungsmaßes auf dem Identifikationsdokument (303) aufweist, und  
wobei der Prozessor (103) ausgebildet ist, das Identifikationsdokument (303) in dem  
Dokumentenbild (305) auf Basis des erfassten Textbereiches (301) und des  
10 vorbestimmten Textbereichsanordnungsmaßes zu erfassen.

12. Mobilgerät (100) nach einem der vorstehenden Ansprüche, wobei der Prozessor  
(103) ausgebildet ist, eine Raumlage der Bildkamera (101) relativ zu dem erfassten  
Textbereich (301) zu bestimmen, wobei die Bildkamera (101) ausgebildet ist, ein weiteres  
15 Bild des Identifikationsdokumentes (303) zu erfassen, um ein weiteres Dokumentenbild zu  
erhalten, und wobei der Prozessor (103) ausgebildet ist, den Textbereich (301) in dem  
weiteren Dokumentenbild unter Verwendung der bestimmten Raumlage zu erfassen.

13. Mobilgerät (100) nach einem der vorstehenden Ansprüche, wobei der Textbereich  
20 (301) eine maschinen-lesbare Zone des Identifikationsdokumentes (303) ist.

14. Verfahren (200) zum Erfassen eines Textbereiches (301) auf einem  
Identifikationsdokument (303), wobei der Textbereich (301) eine Mehrzahl von  
Textzeichen in einer vorbestimmten Anordnung gemäß einem vorbestimmten  
25 Anordnungsmaß aufweist, mit:

Erfassen (201) eines Bildes des Identifikationsdokumentes (303), um ein Dokumentenbild  
(305) zu erhalten;

30 Segmentieren (203) des Dokumentenbildes (305), um eine Mehrzahl von Bildsegmenten  
zu erhalten;

Auswählen (205) einer Mehrzahl von Textzeichenbildsegmenten aus der Mehrzahl von  
Bildsegmenten, wobei die Textzeichenbildsegmente jeweils ein Textzeichen  
35 repräsentieren;

Bestimmen (207) einer Mehrzahl von Textzeichengruppen auf Basis der Mehrzahl von  
Textzeichenbildsegmenten, wobei die Textzeichengruppen jeweils eine Folge von  
Textzeichenbildsegmenten umfassen; und

40

- 5    Vergleichen (209) einer Mehrzahl von Anordnungsmaßen der Mehrzahl von Textzeichengruppen mit dem vorbestimmten Anordnungsmaß, um den Textbereich (301) auf dem Identifikationsdokument (303) zu erfassen.
- 10    15.    Computerprogramm mit einem Programmcode zum Ausführen des Verfahrens (200) nach Anspruch 14, wenn das Computerprogramm auf einem Computer ausgeführt wird.

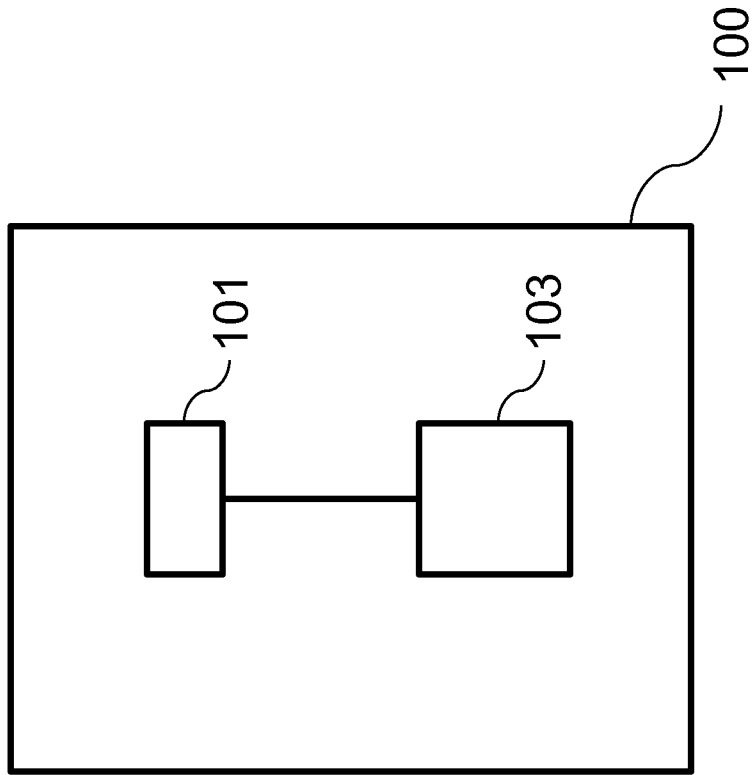


Fig. 1

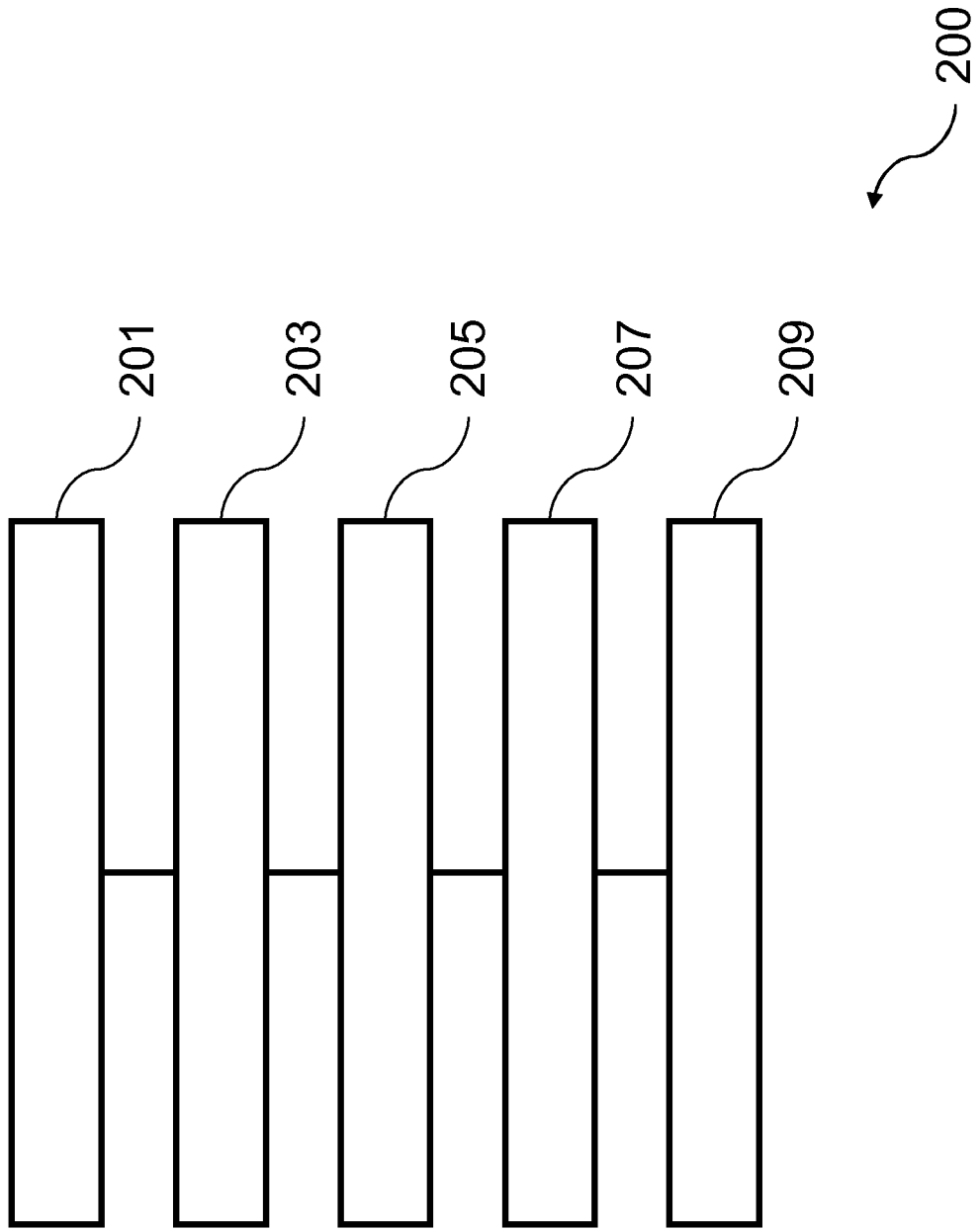


Fig. 2

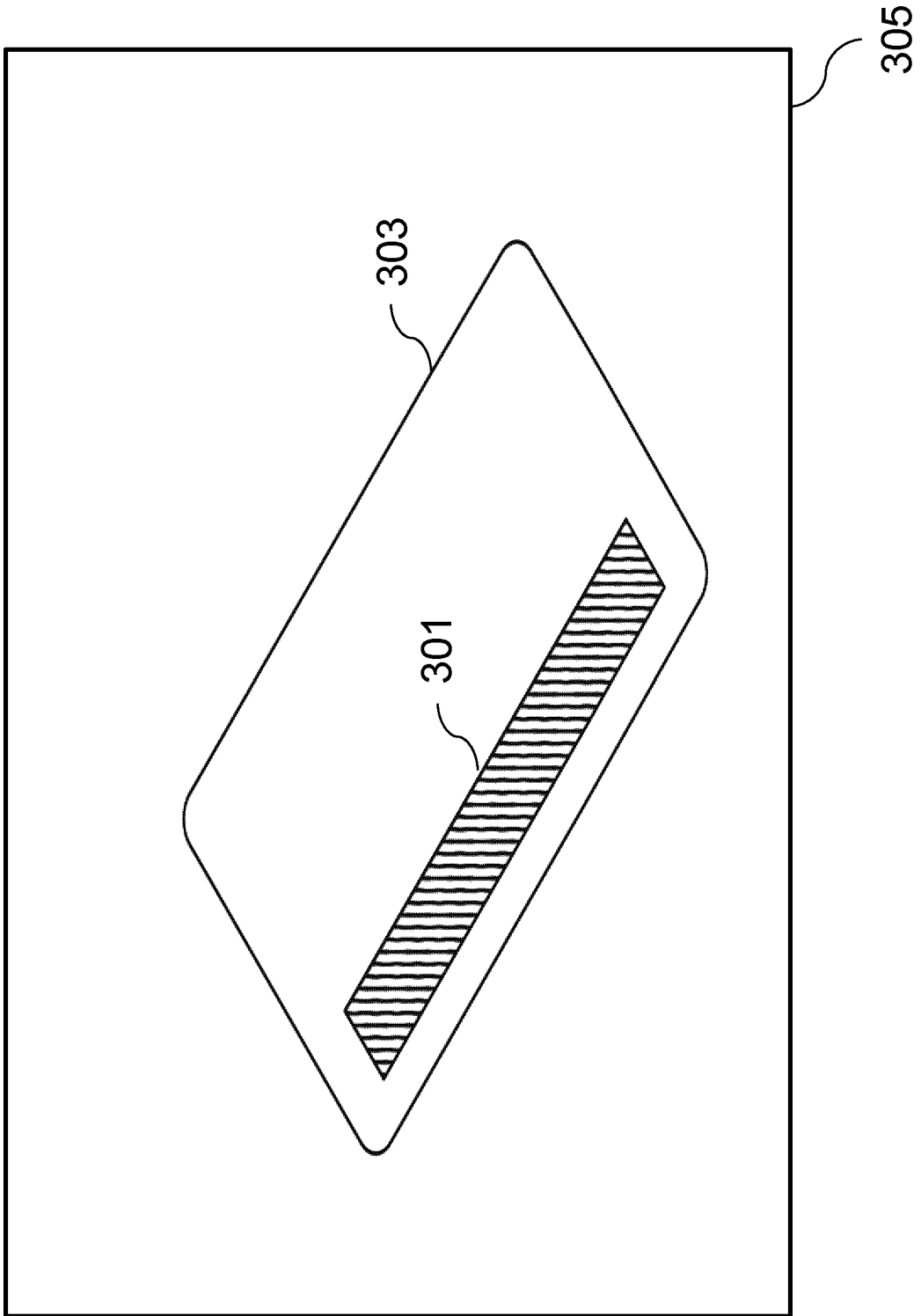


Fig. 3

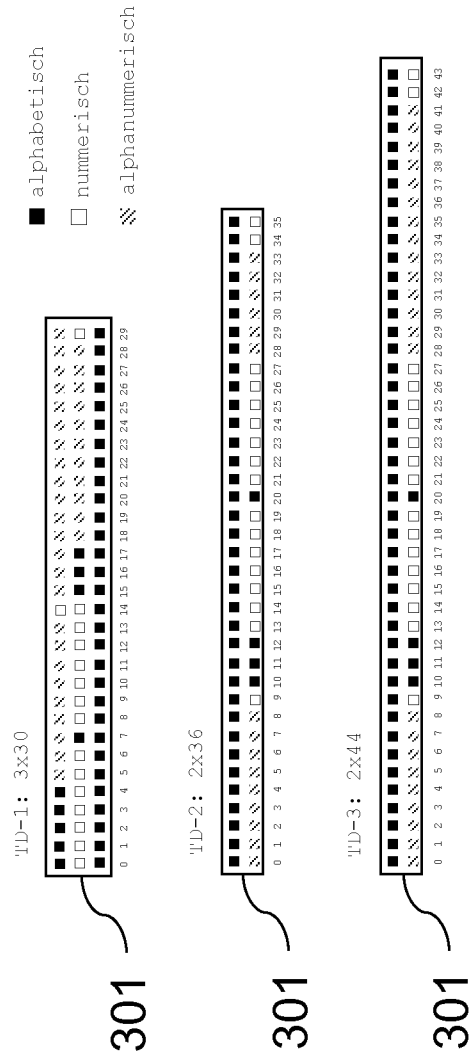


Fig. 4

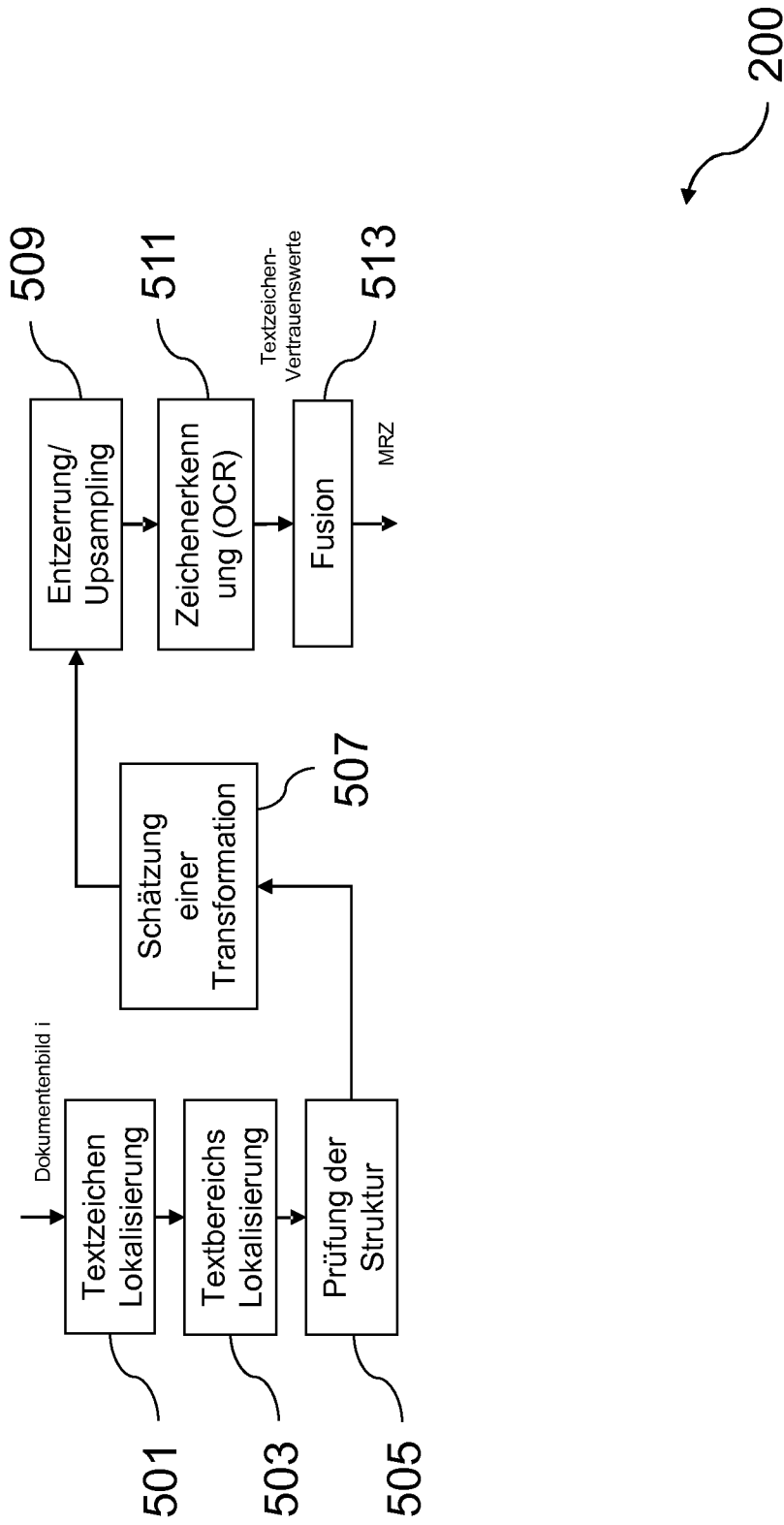


Fig. 5

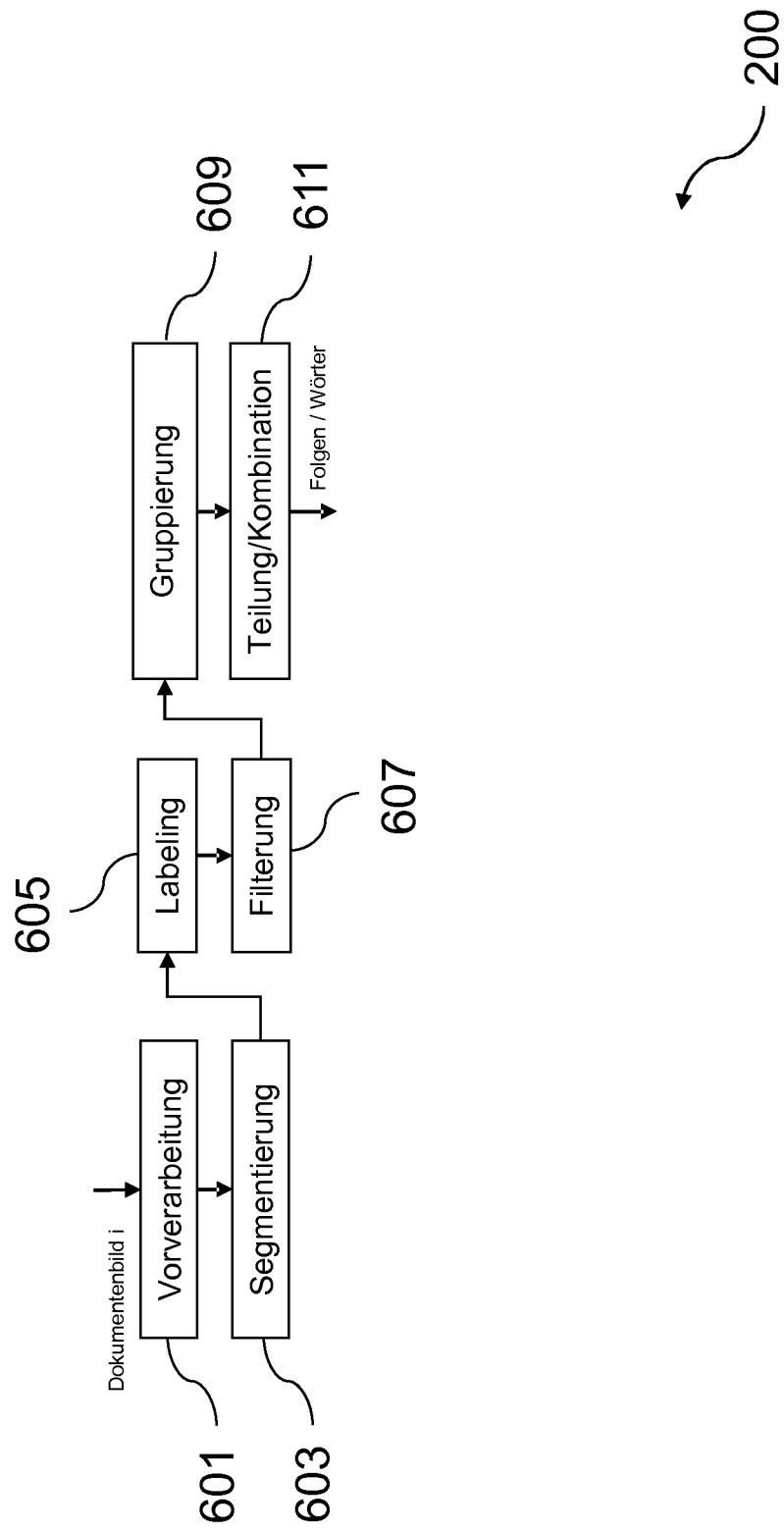


Fig. 6

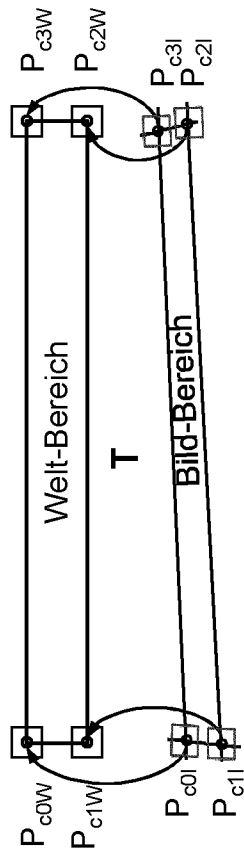


Fig. 7

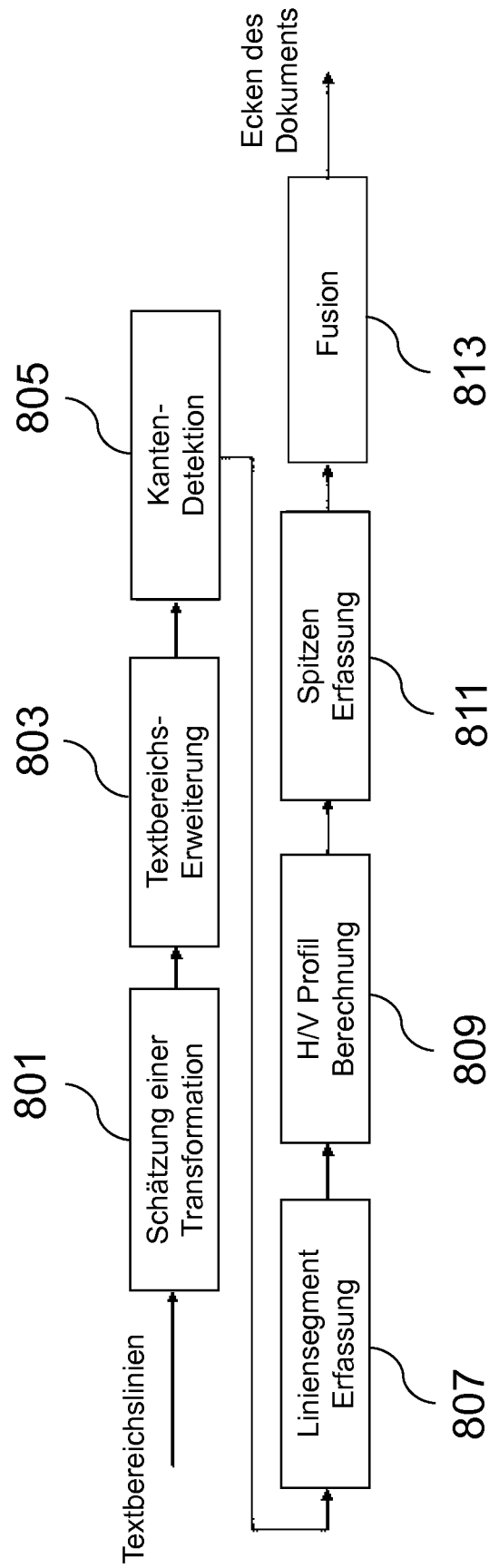


Fig. 8

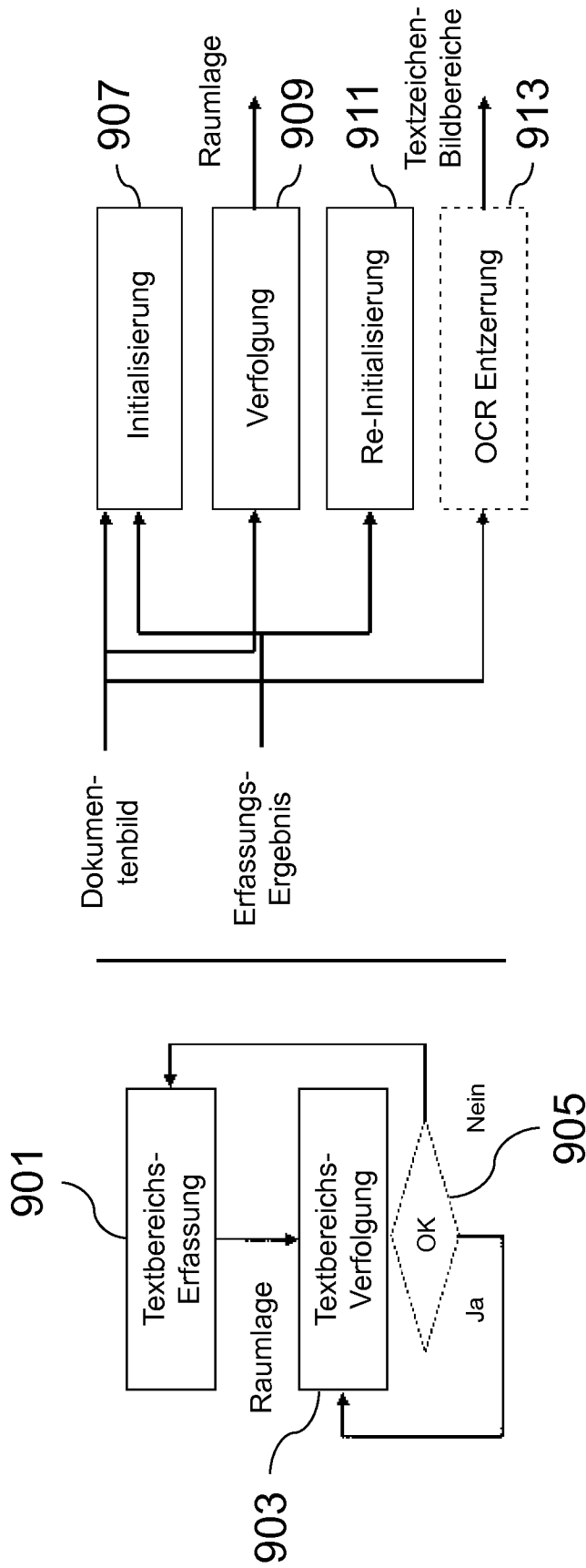


Fig. 9

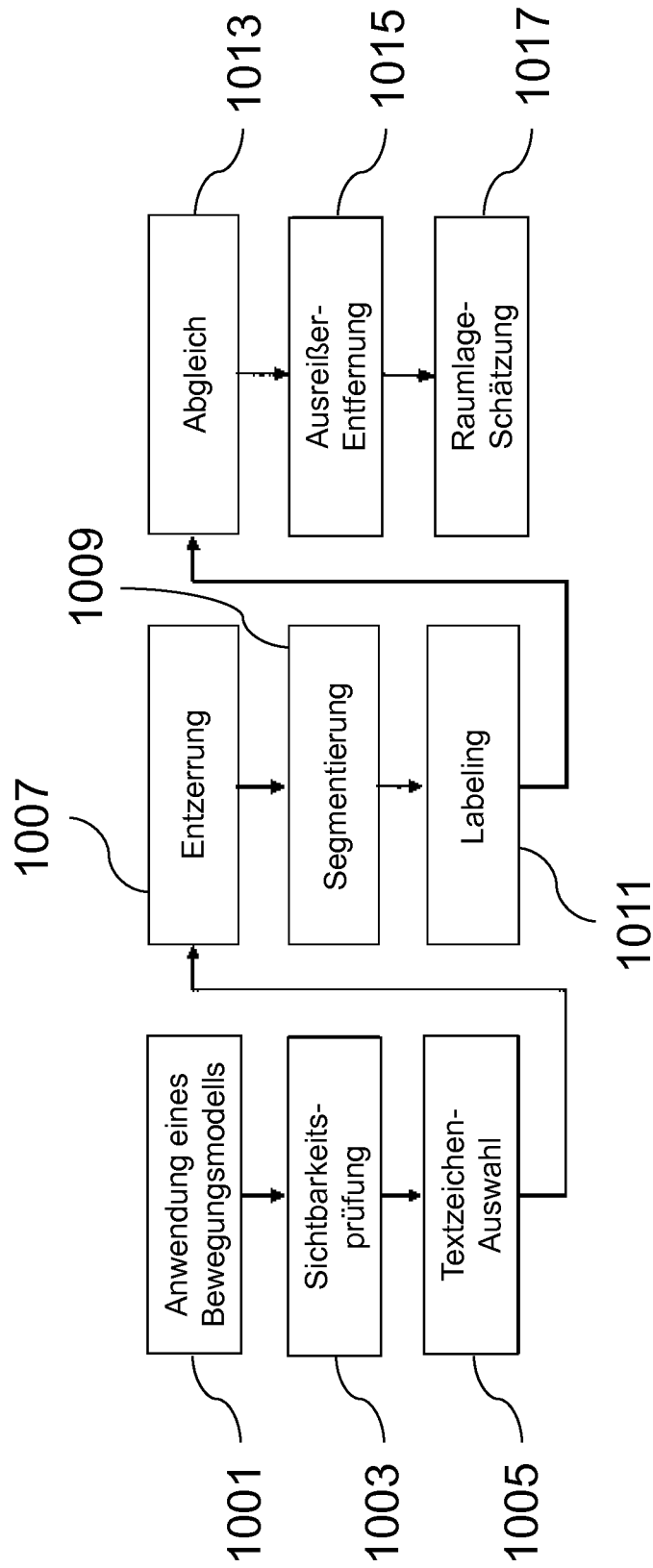


Fig. 10

INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2016/053242

A. CLASSIFICATION OF SUBJECT MATTER  
INV. G06K9/00 G06K9/32  
ADD.  
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED  
Minimum documentation searched (classification system followed by classification symbols)  
G06K  
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 2 320 390 A1 (ICAR VISION SYSTEMS SL [ES]) 11 May 2011 (2011-05-11)	1,3,6-15
Y	paragraphs [0001], [0013], [0014], [0026], [0027], [0034], [0041]; claim 1; figures 1, 2a, 2b	2,4,5
Y	FABRIZIO J ET AL: "Text segmentation in natural scenes using Toggle-Mapping", IMAGE PROCESSING (ICIP), 2009 16TH IEEE INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 7 November 2009 (2009-11-07), pages 2373-2376, XP031628154, ISBN: 978-1-4244-5653-6 the whole document	2

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search 12 May 2016	Date of mailing of the international search report 20/05/2016
--	--

Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Nowbakht Irani, Ali
--	---

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2016/053242

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>RAZAFINDRAMANANA OCTAVIO ET AL: "Alpha*-Approximated Delaunay Triangulation Based Descriptors for Handwritten Character Recognition", 2013 12TH INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION, IEEE, 25 August 2013 (2013-08-25), pages 440-444, XP032502761, ISSN: 1520-5363, DOI: 10.1109/ICDAR.2013.95 [retrieved on 2013-10-10] the whole document -----</p>	4,5

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2016/053242

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
EP 2320390	A1	11-05-2011	CA 2779946 A1	19-05-2011
			CO 6541544 A2	16-10-2012
			EP 2320390 A1	11-05-2011
			US 2012281077 A1	08-11-2012
			WO 2011058418 A2	19-05-2011
-----				

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES  
 INV. G06K9/00 G06K9/32  
 ADD.  
 Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC

B. RECHERCHIERTE GEBIETE  
 Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole )  
 G06K

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

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)  
 EPO-Internal, WPI Data

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	EP 2 320 390 A1 (ICAR VISION SYSTEMS SL [ES]) 11. Mai 2011 (2011-05-11)	1,3,6-15
Y	Absätze [0001], [0013], [0014], [0026], [0027], [0034], [0041]; Anspruch 1; Abbildungen 1, 2a, 2b	2,4,5
Y	FABRIZIO J ET AL: "Text segmentation in natural scenes using Toggle-Mapping", IMAGE PROCESSING (ICIP), 2009 16TH IEEE INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 7. November 2009 (2009-11-07), Seiten 2373-2376, XP031628154, ISBN: 978-1-4244-5653-6 das ganze Dokument	2

Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen  Siehe Anhang Patentfamilie

\* Besondere Kategorien von angegebenen Veröffentlichungen :

- "A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
- "E" frühere Anmeldung oder Patent, die bzw. das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
- "L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
- "O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
- "P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist
- "T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist
- "X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden
- "Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist
- "&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche	Absenddatum des internationalen Recherchenberichts
12. Mai 2016	20/05/2016

Name und Postanschrift der Internationalen Recherchenbehörde Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Bevollmächtigter Bediensteter  Nowbakht Irani, Ali
--	--

C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
Y	<p>RAZAFINDRAMANANA OCTAVIO ET AL: "Alpha*-Approximated Delaunay Triangulation Based Descriptors for Handwritten Character Recognition", 2013 12TH INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION, IEEE, 25. August 2013 (2013-08-25), Seiten 440-444, XP032502761, ISSN: 1520-5363, DOI: 10.1109/ICDAR.2013.95 [gefunden am 2013-10-10] das ganze Dokument -----</p>	4,5

**INTERNATIONALER RECHERCHENBERICHT**

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/EP2016/053242

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
EP 2320390	A1	11-05-2011	CA 2779946 A1 19-05-2011
			CO 6541544 A2 16-10-2012
			EP 2320390 A1 11-05-2011
			US 2012281077 A1 08-11-2012
			WO 2011058418 A2 19-05-2011
-----			