

SCHWEIZERISCHE EIDGENOSSENSCHAFT
EIDGENÖSSISCHES INSTITUT FÜR GEISTIGES EIGENTUM

(11) **CH** **714 891 A2**

(51) Int. Cl.: **G06T 7/10** (2017.01)
G06N 3/08 (2006.01)

Patentanmeldung für die Schweiz und Liechtenstein

Schweizerisch-liechtensteinischer Patentschutzvertrag vom 22. Dezember 1978

(12) **PATENTANMELDUNG**

(21) Anmeldenummer: 00502/18

(71) Anmelder:
Parashift AG, Gewerbstrasse 12
4450 Sissach (CH)

(22) Anmeldedatum: 13.04.2018

(43) Anmeldung veröffentlicht: 15.10.2019

(72) Erfinder:
Andre Bieler, 4103 Bottmingen (CH)
Alain Veuve, 4450 Sissach (CH)

(54) **Bildbearbeitung unter Verwendung von neuronalen Netzwerken.**

(57) Anhand von synthetisch erzeugten Beispieldaten wird ein neuronales Netzwerk darauf hin trainiert, für jedes Pixel in einem potentiell verunreinigten Bild eine Vorhersage für dessen korrigierten Graustufenwert zu machen. Das neu entstandene Bild aus berechneten Graustufenwerten hat schliesslich bessere Eigenschaften für die gängigen Texterkennungsprogramme als das Ursprungsbild.

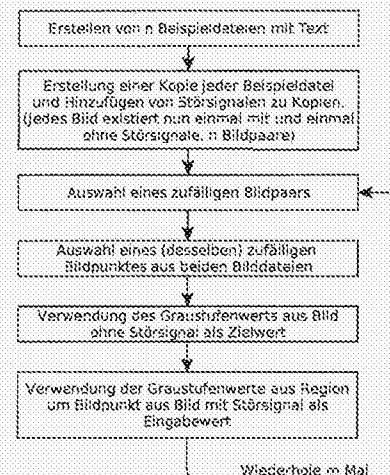


Abbildung 1: Diagramm zur Erzeugung von Trainingsdaten.

Beschreibung

Technisches Gebiet

[0001] Die vorliegende Erfindung bezieht sich allgemein auf ein Verfahren zur automatischen Aufbesserung von Textinformationen in Bilddateien und fällt in das Gebiet der digitalen Bildbearbeitung und optischen Texterkennung.

Stand der Technik

[0002] Optical Character Recognition (OCR, Optische Zeichenerkennung) ist eine Technologie die bei der Digitalisierung von Dokumenten weite Verbreitung findet. Dabei wird versucht aus Bilddokumenten Textzeichen zu extrahieren, welche dann in digitaler Form weiter verwendet werden können.

[0003] Die Erkennungsrate und Genauigkeit der Textextraktion ist dabei stark von der Bildqualität abhängig. Eine Vorverarbeitung der zu untersuchenden Bildern ist ein zentraler Verarbeitungsschritt, mit dem Ziel die Qualität der Textextraktion möglichst positiv zu beeinflussen. Es gilt dabei, qualitätsmindernde Merkmale im Bild zu entfernen, ohne dabei textliche Inhalte zu verlieren.

[0004] Die verwendeten Methoden sind zahlreich und können untereinander kombiniert, oder verkettet werden. Als Beispiele sollen hier zwei Methodengruppen genannt werden: 1) Tief- und Hochpassfilter-Methoden: Diese werden zur Schärfung oder Glättung von Bildern verwendet. Tiefpassfilter reduzieren dabei kleine Störsignale durch Mittelung der Werte über einen fest definierten Bereich. Im Gegenzug werden gleichzeitig scharfe Kanten verwaschen. Umgekehrt sind Hochpassfilter dazu geeignet Kanten zu betonen, verstärken aber auch den Effekt kleinerer Verunreinigungen. 2) Schwellwertmethoden: Jedem Bildpunkt wird einer von zwei Werten zugewiesen, je nachdem ob der Bildpunkt den vorher definierten Schwellwert über- oder unterschreitet. Bei Graustufenbildern resultiert dabei ein schwarz-weisses Binärbild. Es wird zwischen globalen und adaptiven Schwellwertmethoden unterschieden, bei welchen der Schwellwert entweder über das gesamte Bild angewendet wird, oder ob sich dieser Schwellwert dynamisch an verschiedene Bildregionen anpasst.

Nachteil

[0005] Je nach Art der Bildverunreinigungen müssen mehrere Prozessschritte ausgeführt werden, die einzeln von spezifischen Parametern abhängig sind (z.B. die Grösse des Mittelungsbereichs bei Tiefpassfiltermethoden oder der Schwellwert selbst). Dabei gibt es kein Standardverfahren um optimale Ergebnisse zu finden. Die verschiedenen Methoden stehen oft in einem Zielkonflikt indem sich ein gewisser Aspekt der Bildqualität verbessert, dafür andere negative Effekte verstärkt werden. So sind z.B. bei der Rauschunterdrückung durch Tiefpassfilter die entstehenden weichen Kanten oft unerwünscht.

Aufgabe/Vorteile

[0006] Patentanspruch 1 ist eine weitere Methode digitaler Bildbearbeitung zur Verbesserung von OCR-Ergebnissen. Dabei wird ein oder mehrere Neuronale Netzwerke darauf trainiert diese Aufgabe zu erfüllen. Es gibt bei der Anwendung keine Parameter festzulegen und es besteht kein intrinsischer Zielkonflikt zwischen Text-hervorhebenden- und Störeffekte-unterdrückenden Massnahmen.

[0007] Aufzählung der Figuren

Abbildung 1 zeigt eine grafische Darstellung der Methode zur Erzeugung von Trainingsdaten die zur Erstellung des Neuronalen Netzwerks benötigt werden. Beim Erstellen der Beispieldaten ist darauf zu achten, dass möglichst realistische Dokumente erzeugt werden. Dies in Bezug auf die textlichen Inhalte, sowie in der Gliederung (Man sollte z.B. nicht einfach zufällige Buchstabenreihenfolgen wählen.). Im Weiteren sollen verschiedene Schriftarten und -grössen verwendet werden.

Abbildung 2 zeigt einen typischen Ausschnitt einer Beispieldatei mit Störsignalen. Um den korrigierten Graustufenwert des schwarz umrandeten Bildpunktes zu berechnen, werden die Bildpunkte rund um diesen Punkt in Betracht gezogen (umliegendes schwarz umrandetes Quadrat). Die Bildpunkte dieser Fläche aus der Bilddatei mit Störsignal wird als Eingabe in den Trainingsdaten verwendet, der Graustufenwert des eingerahmten Bildpunktes aus der Bilddatei ohne Störsignale als vorgegebener Ausgabewert (= label).

Abbildung 3 zeigt ein Beispiel für ein Label-Datei (ohne Störsignale) 1a) und Daten-Datei (mit Störsignalen) 1b) Paar. Beide Dateien haben dasselbe Format.

Beschrieb

[0008] Neuronale Netzwerke werden anhand von Beispieldaten trainiert um zu einem späteren Zeitpunkt Entscheidungen und Vorhersagen aufgrund neuer Daten zu treffen. Im Fall von Patentanspruch 1 besteht ein solcher Datensatz aus einer Ansammlung von Bildern, die Text enthalten.

Erstellen von Beispieldateien

[0009] Beispiel-/Trainings-Dateien bestehen aus Daten-Dateien und Label-Dateien. Daten-Dateien sind digitale Bilder in einem Pixelgrafik-Format wie .jpeg, .png oder ähnliches und können ein beliebiges Format aufweisen. Die Label-Dateien enthalten die korrekten Graustufenwerte für jedes Pixel der Daten-Dateien. Zur Herstellung von synthetischen Trainingsdatensätzen werden erst Label-Dateien erstellt. Dies geschieht durch ein beliebiges Textverarbeitungsprogramm, sofern dieses den Export als Bitmap oder pdf unterstützen. Die Daten-Dateien sind zunächst Kopien der Label-Dateien. Ihnen werden in einem weiteren Schritt aber noch Störeffekte hinzugefügt. Ein Beispiel für ein Paar aus Label- und Daten-Datei ist in Abbildung 3a) (Label) und Abbildung 3b) (Daten) zu sehen. Es sollen möglichst verschiedene (gängige) Schrift-Arten und Grössen verwendet werden um bei der Textsäuberung später gute Resultate zu erzielen. Ebenfalls wichtig ist es, möglichst realistische Störeffekte einzuarbeiten.

Erstellen von Trainingsdatensätzen

[0010] Ein Trainingsdatensatz besteht aus Daten ($d_1 \dots d_m$) und Labeln ($l_1 \dots l_m$). Diese werden aus Daten-Dateien und Label-Dateien gewonnen. Die eigentlichen Daten sind Ausschnitte aus Daten-Dateien in einem Format von $n_x \times n_y$ Pixeln. Wobei wir $n_x = n_y = 129$ verwenden. n_x und n_y müssen nicht identisch, aber beide ungerade sein, damit es einen Zentralen Bildpunkt gibt. Das dazugehörige Label entspricht dem Pixelwert des zentralen Pixels aus demselben Ausschnitt der Label-Datei (siehe Abbildung 2).

[0011] Es gibt n Daten-Dateien ($D^{(1)} \dots D^{(n)}$) und n Label-Dateien ($L^{(1)} \dots L^{(n)}$). Die zu jeder Daten-Datei zugehörige Label-Datei hat dasselbe Format in Pixeln wie die Daten-Datei. (Format $L^{(1)} = \text{Format } D^{(1)}$) Daten-Dateien können unterschiedliche Formate haben (Format $D^{(i)} \neq D^{(j)}$). Folgender Algorithmus generiert dann Trainingsdatensätze:

1. Wähle zufällig eine der n Label-Dateien $L^{(i)}$ und die dazugehörige Daten-Datei $D^{(i)}$.
2. Wähle zufällig die Koordinaten $p = (x_j^{(i)}, y_k^{(i)})$ eines Bildpunktes aus $L^{(i)}$.
3. Das Label l ist dann der Wert des Bildpunktes $(x_j^{(i)}, y_k^{(i)})$ in $L^{(i)}$.
4. Das Datum d ist dann der 129×129 Pixel Bildausschnitt um den Bildpunkt $(x_j^{(i)}, y_k^{(i)})$ aus $D^{(i)}$, ausschliesslich des Zentralen Bildpunktes, für den es den Graustufenwert zu berechnen gilt. Das heisst, es wird in $D^{(i)}$ von p aus ein Ausschnitt von 64 Pixeln nach links, rechts, oben und unten genommen.
5. Wiederhole 1. bis 4. m -Mal

Erstellung des neuronalen Netzwerks

[0012] Das spezifische Design des Neuronen Netzwerks soll hier nicht vorgegeben werden, da dies je nach Anwendungsfall, verfügbarer Rechner-Ressourcen und Menge an Trainingsdaten stark variieren kann.

Anwendung des Neuronen Netzwerks

[0013] Hat man nun eine neue Bilddatei B , wird für jedes Pixel (n_x, n_y) , (mit $64 < n_x < N_x - 64$ und $64 < n_y < N_y - 64$), eine Vorhersage des Neuronen Netzwerks berechnet. Wobei N_x die totale Anzahl Pixel in x-Richtung und N_y die totale Anzahl Pixel in y-Richtung von B sind. Somit entsteht ein neues, gesäubertes Bild vom Format $(N_x - 128 \times N_y - 128)$. Das Ausgangsbild wird dabei nicht Pixel für Pixel durch die Ausgabe des Neuronen Netzwerks ersetzt, sondern es entsteht ein neues Bild ausschliesslich aus den berechneten Graustufenwerten.

Patentansprüche

1. Verfahren zur automatischen Verbesserung von Textinformation in Bilddateien, gekennzeichnet durch die Verwendung von Neuronen Netzwerken.
2. Verfahren zur automatischen Verbesserung von Textinformationen in Bilddateien nach Anspruch 1, gekennzeichnet durch die Berechnung korrigierter Graustufenwerte sämtlicher Bildpunkte.
3. Verfahren zur automatischen Verbesserung von Textinformationen in Bilddateien nach Anspruch 2, gekennzeichnet dadurch, dass zur Berechnung der neuen Graustufenwerte jeweils ein quadratischer Bildausschnitt von 129×129 Pixeln um den Bildpunkt verwendet wird.
4. Verfahren zur automatischen Verbesserung von Textinformationen in Bilddateien nach Anspruch 3, gekennzeichnet durch die Verwendung von Neuronen Netzwerken zur Berechnung der korrigierten Graustufenwerte.
5. Verfahren zur automatischen Verbesserung von Textinformationen in Bilddateien nach Anspruch 4, gekennzeichnet durch die Verwendung von Neuronen Netzwerken welche mit synthetisch erzeugten Beispieldaten nach Abbildung 1 trainiert wurde.

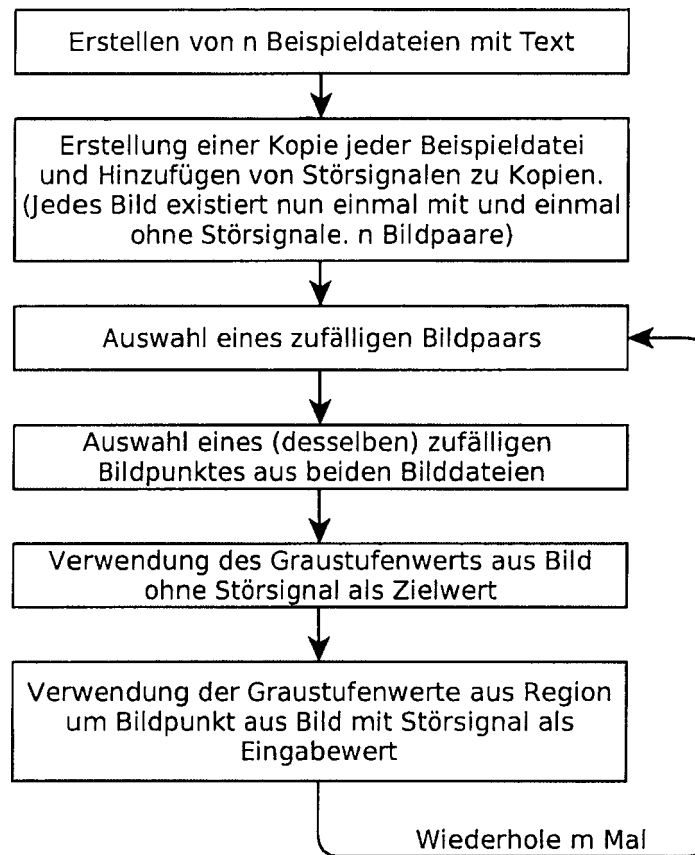


Abbildung 1: Diagramm zur Erzeugung von Trainingsdaten.

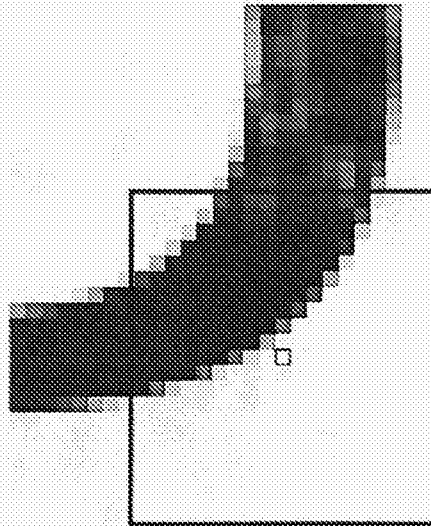
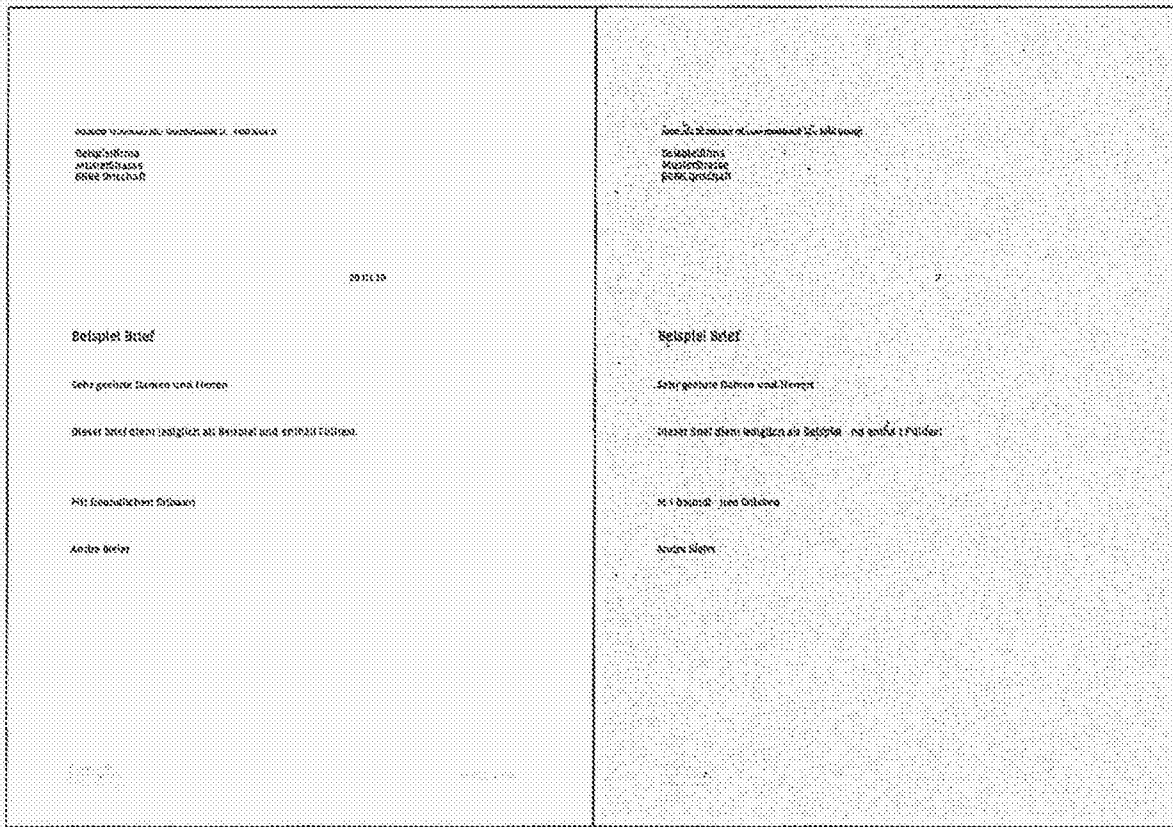


Abbildung 2: Zur Berechnung des Graustufenwertes des schwarz umrandeten Pixel werden die Werte der umliegenden Zone (grösseres schwarzes Quadrat) berücksichtigt. Zu Illustrationszwecken ist die umliegende Fläche kleiner als $129 \text{ Pixel} \times 129 \text{ Pixel}$ gehalten.



(a) Beispiel einer Label Datei

(b) Beispiel Daten Datei

Abbildung 3