

Entzerrung von Textdokumenten unter Verwendung von Thin-Plate-Spline

Marco Block und Raúl Rojas
 Freie Universität Berlin
 Institut für Informatik
 Takustr. 9, 14195 Berlin, Germany
 {block,rojas}@inf.fu-berlin.de

September 2008

Zusammenfassung—Viele Verfahren identifizieren zunächst Textzeilen und ermitteln anschließend über die Dokumentrandeigenschaften die notwendige Korrekturtransformation. Das in dieser Arbeit vorgestellte Verfahren benötigt keine Dokumentrandeigenschaften, sondern ermittelt die Dokumentorientierung anhand der Buchstabenausrichtungen. Als Interpolationmethode wird Thin-Plate-Spline verwendet.

I. EINFÜHRUNG

Viele Texterkennungsverfahren setzen entzerrungsfreie, planare, korrekt rotierte Dokumente voraus. Aufnahmen von planaren Dokumenten, bei denen lediglich die Rotation korrigiert werden soll, lassen sich z.B. mit der Hough-Transformation [13], projektionsbasierte [1], korrelationsbasierte Methoden [12] oder mit der Fourier-Transformation [20] erfolgreich korrigieren.

Einige Methoden machen Annahmen über das geometrische Modell eines Dokuments und passen dieses nach lokalen Eigenschaften der Aufnahme entsprechend an [16].

Im Laufe der letzten Jahre, wurden zahlreiche Ansätze entwickelt und untersucht [15], [7]. Dabei identifizieren einige Ansätze zunächst die Textzeilen und gehen anschließend davon aus, dass der Textfluss in Blockform vorliegt. Anhand dieser Informationen lässt sich ein Raster konstruieren, durch das eine Korrektur des Bildes ermöglicht wird.

Andere Verfahren betten die Bildinformationen in ein Modell ein und passen dieses solange an, bis der entstehende Fehler minimiert wird [26], [11].

II. IDENTIFIKATION DER TEXTZEILEN

A. Merkmalsextraktion

In der ersten Phase werden die Textzeilen identifiziert. Ein Textblock \mathbf{I} sollte dabei relativ gerade gedreht vorliegen. Um die Orientierungen der Textzeilen zu identifizieren, werden über m horizontale, schmale Bildstreifen der Breite w , die lokalen Maxima innerhalb des Kantenbildes $\mathcal{F}^{SY}(\mathbf{I})$ ermittelt, mit $\mathcal{F}^{SY}(\mathbf{I})$ wendet nur den Y -Gradienten des Sobelfilters auf ein binarisiertes Bild \mathbf{I} an. Mit w_k bezeichnen wir den k -ten Bildstreifen, wobei $k \in \{0, 1, \dots, m\}$ ist. Nicht in jedem Fenster lassen sich Maxima identifizieren.

Um die Maxima zu bestimmen, werden für alle Bildstreifen w_i Histogramme h_i senkrecht zur erwarteten Textrichtung

aufgestellt. Histogrammwerte $h(k)$ werden auf Null gesetzt, für die $h(k) \leq \frac{w}{5}$ gilt. Im nächsten Schritt wird der Mittelwert benachbarter, positiver Histogrammwerte auf den entsprechenden Schwerpunkt gesetzt. Im Kantenbild sind neben den erwünschten Kanten unterhalb einer Textzeile, auch Teile der Buchstaben enthalten. Die konkurrierenden lokalen Maxima liegen sehr nahe beieinander und so kann ein direkter Vergleich das gewünschte Maximum filtern.

Die identifizierten Merkmale für ein Textblockbeispiel sind in Abbildung 1 gezeigt.

1. Introduction

Signature verification is a biometric attribute. Well-known biometric methods include iris, retina, face- and fingerprint-based identification and verification [1]. They are being used more and more in our ordinary life. Although attributes like iris, retina and fingerprints do not change over time, they require special and relatively expensive hardware to capture the image. An important advantage of the signature verification compared with other biometric attributes is its long tradition in many common commercial fields [2], such as e-business, which includes online banking transactions, electronic payments, access control and so on. So signature verification is a very popular research area right now. Generally, it is accepted that an individual's signature is unique, although nobody verifies it. In fact signature verification is a difficult pattern problem because the intra-class variations could be large.

Abbildung 1. Identifizierte Textmerkmale über gefilterte Maxima in den vertikalen Histogrammen.

B. Merkmalsverkettung

Die identifizierten Maxima werden zu n Punktelisten L zusammengefasst (siehe Abbildung 2). Um zu entscheiden, welcher Punkt eine Liste L_i an der Stelle $L_i[j] = (x, y)$ fortsetzt, wird der Fehler zwischen den zur Verfügung stehenden Punkten und dem Vorhersagepunkt $P(L_i[j], \varphi_j)$ berechnet

$$P((x, y), \varphi) = (L_i[j].x + w \cdot \cos \varphi, -L_i[j].y + w \cdot \sin \varphi),$$

wobei φ_j der Orientierung an der Position $L_i[j]$ entspricht.

Sollte im Bildstreifen w_{j+1} kein Punkt in Frage kommen, da die oberen, bzw. unteren Linien, diesen Punkt bereits in Anspruch nehmen, wird der Vorhersagepunkt an der Stelle $P(L_i[j], \varphi_j)$ für das nächste Fenster w_{j+2} neu berechnet, usw.

Jede Linie L_i kann dabei eine unterschiedliche Anzahl von m_i Maxima aufweisen. Da sich die Nachbarlinien ober- und unterhalb einer Linie innerhalb eines Dokuments ähnlich verhalten sollten, minimiert die Wahl der n Linienmodelle eines Textblocks den allgemeinen Fehler E_{TB}

$$E_{TB} = \sum_{i=1}^n E_i,$$

wobei E_i den Fehler eines Linienmodells L_i darstellt

$$E_i = \sum_{j=1}^{m_i} \text{dist}(L_i[j], P(L_i[j], \varphi_j)).$$

Die Linienmodelle werden in Abbildung 2 farblich dargestellt.

1. Introduction

Signature verification is a biometric attribute. Well-known biometric methods include iris, retina, face and fingerprint-based identification and verification [1]. They are being used more and more in our ordinary life. Although attributes like iris, retina and fingerprints do not change over time they require special and relatively expensive hardware to capture the image. An important advantage of the signature verification compared with other biometric attributes is its long tradition in many common commercial fields [2], such as e-business, which includes on-line banking transactions, electronic payments, access control and so on. So signature verification is a very popular research area right now. Generally, it is accepted that an individual's signature is unique, although nobody verifies it. In fact signature verification is a difficult pattern problem because the intra-class variations could be large.

Abbildung 2. Die Maxima der Histogramme werden verkettet und sind farblich unterschiedlich dargestellt.

C. Splinerepräsentation der Merkmalslisten

Nachdem n Linienmodelle fest stehen, werden diese durch n geglättete Funktionen (*Splines*) repräsentiert. Es gibt zahlreiche Methoden für die Approximation von Funktionen die eine Kernel-Glättung mit entsprechendem Glättungsfaktor h verwenden [14]. Für die Glättung der Textzeilen wird ein Gaußkernel K verwendet

$$S(x, h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right).$$

Aus den berechneten n Splines werden jeweils mit gleicher Schrittweite, beginnend bei $x = 0$ die Ableitungen der Funktionen ermittelt. Die gefundenen Ableitungen sind nur innerhalb des Start- und Endpunkts einer Funktion definiert (siehe Abbildung 3).

Bis zu diesem Arbeitsschritt ähnelt das Verfahren der Beschreibungen in den Arbeiten [23], [24].

D. Ableitungen im X-Vektorfeld

Jetzt werden die Ableitungen in X-Richtung berechnet. Für die folgende Interpolation der Steigungen wird in den Arbeiten [23], [24] die Delaunay Triangulierung [6] für bekannte

1. Introduction

Signature verification is a biometric attribute. Well-known biometric methods include iris, retina, face and fingerprint-based identification and verification [1]. They are being used more and more in our ordinary life. Although attributes like iris, retina and fingerprints do not change over time they require special and relatively expensive hardware to capture the image. An important advantage of the signature verification compared with other biometric attributes is its long tradition in many common commercial fields [2], such as e-business, which includes on-line banking transactions, electronic payments, access control and so on. So signature verification is a very popular research area right now. Generally, it is accepted that an individual's signature is unique, although nobody verifies it. In fact signature verification is a difficult pattern problem because the intra-class variations could be large.

Abbildung 3. Die identifizierten, langen Merkmalsketten werden durch Splines approximiert.

und eine baryzentrische Interpolation für unbekannte Punkte verwendet. Die baryzentrische Interpolation ermittelt dabei für jeden Pixel p innerhalb eines Dreiecks, das durch die Punkte a, b, c gegeben ist und die bekannten Funktionswerten $f(a)$, $f(b)$ und $f(c)$ besitzt, mit

$$f(p) = \alpha \cdot f(a) + \beta \cdot f(b) + \gamma \cdot f(c)$$

den entsprechenden, gewichteten Mittelwert, wenn $p = \alpha \cdot a + \beta \cdot b + \gamma \cdot c$. Das hat den Nachteil, dass sie nicht glatt ist und deshalb ein kantiges Korrekturbild liefert. Eine glatte Interpolationsfunktion ist dem vorzuziehen.

E. Interpolation mit Thin-Plate-Spline

Um ein vollständiges Mesh für alle Positionen innerhalb des Bildes zu erhalten, werden die fehlenden Ableitungen durch das Verfahren *Thin-Plate-Spline* (TPS) approximiert [8], [27].

Eine Generalisierung der kubischen Splines zu einer 2-dimensionalen Repräsentation von Funktionswerten stellt das TPS-Verfahren dar [8], [4]. Der Name bezieht sich auf eine Analogie in der Physik, bei der eine dünne Blechplatte in Abhängigkeit von einigen bekannten Punkten gebogen wird [18], [9].

Durch gegebene Beispielrepräsentanten im \mathbb{R}^2 findet das TPS-Interpolationsverfahren eine minimal gebogene, glatte Fläche, die diese Punkte enthält. Sind drei Punkte gegeben, so entsteht eine ebene Fläche. Für weniger als drei Punkte ist die Funktion nicht definiert und bei mehr als drei Punkten entsteht typischerweise eine gekrümmte Fläche.

Für das TPS-Modell im 2-dimensionalen Fall sind entsprechend aus zwei Bildern \mathbf{I}_1 und \mathbf{I}_2 zwei Punktemengen M_1 und M_2 gegeben mit

$$M_1 = \{(x_{1,i}, y_{1,i}) | i = 1, 2, \dots, n\} \text{ und} \\ M_2 = \{(x_{2,i}, y_{2,i}) | i = 1, 2, \dots, m\}.$$

Die Abbildungsfunktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ für einen unbekanntem Punkt (x, y) ist definiert durch

$$f(x, y) = \sum_{i=1}^n w_i U(|(x_{1,i}, y_{1,i}) - (x, y)|).$$

Dabei ist w_i eine gewichtete 1×2 Matrix und es gilt $U(r) = r^2 \log(r)$. Dabei ist r die euklidische Distanz zwischen dem unbekanntem Pixel (x, y) und den gegebenen Punkten $(x_{1,i}, y_{1,i})$. Ausführliche Modellbeschreibungen finden sich hier [8], [3], [4].

Die Methode findet in unterschiedlichen 2-dim [3], [5], [21], [19] und 3-dim [10] Anwendungen ihren Einsatz. Auch bei der Entzerrung von Fingerabdrücken wurde TPS erfolgreich eingesetzt [22], [25], [2].

Die Ableitungen werden durch TPS geliefert. Da jetzt die Steigung an einer beliebigen Stelle im Bild angegeben werden kann, werden beginnend an der linken Bildseite, in Abhängigkeit der Ableitungen, Linien durch das Bild geschickt. Abbildung 4 zeigt die Konstruktion der horizontalen Linien.

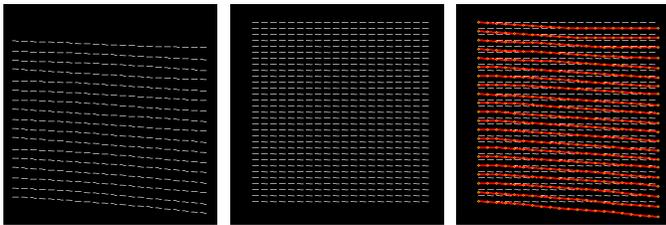


Abbildung 4. Konstruktion der X -Linien anhand der TPS-Interpolation. Splinesableitungen (links), interpolierte Ableitungen (mitte) und neues Linienmodell (rechts).

III. IDENTIFIKATION DER BUCHSTABENORIENTIERUNGEN

Dieser Ansatz identifiziert nicht die Randeigenschaften des Textblocks oder Dokuments, sondern orientiert sich an den Ausrichtungen der Buchstaben. In der Arbeit von Lu, Chen und Ko [17] werden die sogenannte VSBs (*vertical stroke boundary*) einer Textzeile, mit Hilfe morphologischer Operatoren lokalisiert. Der Textzeilenkörper ist in dem dort vorgestellten Verfahren bereits identifiziert und wenige Buchstabenmerkmale genügen, um die Textrichtung anzugeben.

A. Lokale PCA und Clustering

Auf Grund der Tatsache, dass die Identifizierung der Textzeilen mit Hilfe der lokalen PCA stattgefunden hat, und keinerlei Informationen über den ganzen Textkörper vorliegen, muss eine andere Möglichkeit gefunden werden, die Orientierungen der Buchstaben zu finden.

Lokale PCA wird auf das Blockbild $\mathcal{B}^{SB}(\mathbf{I})$ angewendet und diese anschließend geclustert. Dabei kann berücksichtigt werden, dass Buchstaben nicht größer als der durchschnittliche Funktionsabstand sind und nur Orientierungen zwischen $135 - 45^\circ$ relativ zur Zeilenorientierung (gegen den Uhrzeigersinn) erlaubt sind. Abbildung 5 zeigt die resultierenden Cluster der lokalen PCA.

Die Hauptrichtungen der Buchstaben sind dabei gut zu erkennen.

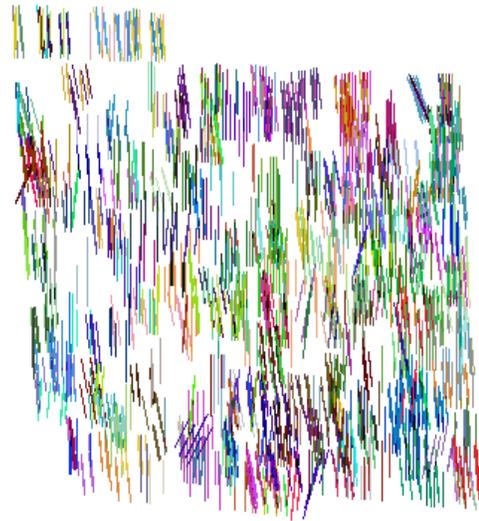


Abbildung 5. Die lokale PCA liefert die Orientierungen der Buchstaben in Y -Richtung.

B. Ableitungen im Y -Vektorfeld mit TPS

Diese Orientierungen werden wieder durch TPS interpoliert. Anschließend werden Linien, diesmal von oben nach unten, durch die interpolierten Daten geschickt. Das Ergebnis ist ein Mesh (siehe 6), deren Schnittpunkte es zu finden gilt.

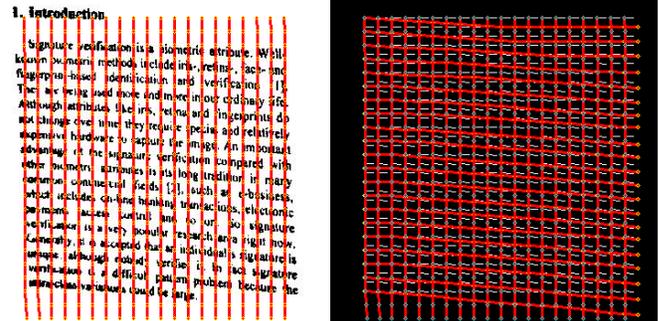


Abbildung 6. Interpolation der Y -Steigungen (links) und Erstellung des kompletten Meshs (rechts).

IV. ENTZERRUNG DES RASTERS

Das Mesh lässt sich jetzt beispielsweise über lokale affine Transformationen mit bikubischer Interpolation entzerren. Ein Ursprungsbild und die dazugehörigen entzerrte Versionen zeigt Abbildung 7.

Dokumentbeispiele, in denen dieses Verfahren schlechtere Ergebnisse als die Standardverfahren liefert, besitzen überwiegend kursive Textpassagen. Dieses Problem kann nur mit einem neuen Verfahren gelöst werden, dass die Textzeilen separat korrigiert.

LITERATUR

- [1] Baird H.S.: „The skew angle of printed documents“, in Proc. SPSE 40th Conf. Symp. Hybrid Imaging Systems, pp.21-24, 1987

1. Introduction

Signature verification is a biometric attribute. Well-known biometric methods include iris-, retina-, face- and fingerprint-based identification and verification [1]. They are being used more and more in our ordinary life. Although attributes like iris, retina and fingerprints do not change over time, they require special and relatively expensive hardware to capture the image. An important advantage of the signature verification compared with other biometric attributes is its long tradition in many common commercial fields [2], such as e-business, which includes on-line banking transactions, electronic payments, access control and so on. So signature verification is a very popular research area right now. Generally, it is accepted that an individual's signature is unique, although nobody verifies it. In fact signature verification is a difficult pattern problem because the inter-class variations could be large.

1. Introduction

Signature verification is a biometric attribute. Well-known biometric methods include iris-, retina-, face- and fingerprint-based identification and verification [1]. They are being used more and more in our ordinary life. Although attributes like iris, retina and fingerprints do not change over time, they require special and relatively expensive hardware to capture the image. An important advantage of the signature verification compared with other biometric attributes is its long tradition in many common commercial fields [2], such as e-business, which includes on-line banking transactions, electronic payments, access control and so on. So signature verification is a very popular research area right now. Generally, it is accepted that an individual's signature is unique, although nobody verifies it. In fact signature verification is a difficult pattern problem because the inter-class variations could be large.

Abbildung 7. Beispiel eines verzerrten und entsprechend entzerrten Textblockes.

- [2] Bazen A.M., Gerez S.H.: „Fingerprint Matching by Thin-Plate Spline Modelling of Elastic Deformations“, Pattern Recognition, Vol. 36, No. 8, pp. 1859-1867, 2003
- [3] Belongie S., Malik J., Puzacha J.: „Matching shapes“, in Proc. 8th Int'l Conf. Computer Vision, Vol.1, pp. 454-461, 2001
- [4] Bookstein F.L.: „Principal Warps: Thin-Plate Splines and the Decomposition of Deformations“, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 11, No.6, pp. 567-585, 1989
- [5] Chui H., Rangarajan A.: „A new algorithm for non-rigid point matching“, in Proc. IEEE Conf. Comput. Vision and Pattern Recognition, Vol.2, pp. 44-51, 2000
- [6] Delaunay B.: „Sur la sphère vide“, Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk, 7:793-800, 1934
- [7] Doermann D., Liang J., Li H.: „Progress in Camera-Based Document Image Analysis“, Seventh International Conference on Document Analysis and Recognition (ICDAR'03), Vol. 1, pp.606-616, 2003
- [8] Donato G., Belongie S.: „Approximation Methods for Thin Plate Spline Mappings and Principal Warps“, ECCV, Vol. 2, pp. 531-542, Kopenhagen/Dänemark, 2002
- [9] Duchon J.: „Splines minimizing rotation invariant seminorms in sobolev spaces“, in Constructive Theory of Functions of Several Variables, Lecture Notes in Mathematics 571 (W. Schempp, K. Zeller, eds.), pp. 85-100, Springer-Verlag, 1977
- [10] Evans A.C., Dai W., Collins D.L., Neelin P., Marrett S.: „Warping of a Computerized 3-D Atlas to Match Brain Image Volumes for Quantitative Neuroanatomical and Functional Analysis“, Proc. SPIE Medical Imaging V: Image Processing, M.H. Loew, ed., Vol.1445, pp.236-246, 1991
- [11] Fu B., Wu M., Li R., Xu Z.: „A model-based book dewarping method using text line detection“, 2nd Int. Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2007), Brasilien, 2007
- [12] Gatos B., Papamarkos N., Chamzas C.: „Skew Detection and Text Line Position Determination in Digitized Documents“, Pattern Recognition, Vol.30, No.9, pp.1505-1519, 1997
- [13] Hinds S.C., Fisher J.L., D'Amato D.P.: „A document skew detection method using run-length encoding and the Hough transform“, in Proc. 10th Int. Conf. on Pattern Recognition, pp. 464-468, 1990
- [14] Jones M.C., Wand M.P.: „Kernel Smoothing (CRC Monographs on Statistics & Applied Probability)“, ISBN 978-0412552700, Chapman&Hall, 2000
- [15] Liang J., Doermann D., Li H.: „Camera-Based Analysis of Text and Documents: A Survey“, International Journal on Document Analysis and Recognition (IJ DAR 2005), 7(2+3), pp. 83-104, 2005
- [16] Lu S., Tan C.L.: „Document Flattening through Grid Modeling and Regularization“, 18th International Conference on Pattern Recognition (ICPR), pp.971-974, Hong Kong, 2006
- [17] Lu S., Chen B.M., Ko C.C.: „Perspective rectification of document images using fuzzy set and morphological operations“, Image and Vision Computing, Vol.23, pp.541-553, 2005
- [18] Meinguet J.: „Multivariate interpolation at arbitrary points made simple“, Journal of Applied Mathematics and Physics (ZAMP) Vol.30, No.2, pp.292-304, 1979
- [19] Powell M.J.: „A thin plate spline method for mapping curves into curves in two dimensions“, in Computational Techniques and Applications: CTAC95, R.L. May and A.K. Easton (Eds.), World Scientific: Singapore 1996, pp. 43-57, 1995
- [20] Pstl W.: „Detection of linear oblique structures and skew scan in digitized documents“, in Proc. 8th Int. Conf. on Pattern Recognition, pp.687-689, 1986
- [21] Rohr K., Fornefett M., Stiehl H.S.: „Approximating Thin-Plate Splines for Elastic Registration: Integration of Landmark Errors and Orientation Attributes“, Proc. 16th Int'l Conf. Information Processing in Medical Imaging (IPMI'99), pp.252-265, 1999
- [22] Ross A.A., Nadgir R.: „A Thin-Plate Spline Calibration Model for Fingerprint Sensor Interoperability“, IEEE Transactions on Knowledge and Data Engineering, Vol. 20, No. 8, pp.1097-1110, 2008
- [23] Schneider D.C., Block M., Rojas R.: „Robust Document Warping with Interpolated Vector Fields“, 9th International Conference on Document Analysis and Recognition (ICDAR 2007), IEEE Computer Society 2007, ISBN 978-0-7695-2822-9, Curitiba/Brazil, 2007
- [24] Schneider D.C.: „Correcting curvature and perspective for camera based document processing“, Masterarbeit, Freie Universität Berlin, 2006
- [25] Shah S., Ross A., Shah J., Crialhameanu S.: „Fingerprint Mosaicing Using Thin Plate Splines“, Biometric Consortium Conference, Arlington/USA, 2005
- [26] Zhang L., Tan C.L.: „Warped document image restoration using Shape-from-Shading and physically-based modeling“, IEEE Workshop on Applications of Computer Vision (WACV), 2007
- [27] Thin Plate Spline editor: <http://elonen.iki.fi/code/tpsdemo/>