

Automatische Qualitätsmessung von Retina-Fundusbildern

Jan Paulus¹, Jörg Meier¹, Rüdiger Bock¹, Joachim Hornegger¹,
Georg Michelson²

¹Lehrstuhl für Informatik 5 (Mustererkennung), Universität Erlangen-Nürnberg

²Augenklinik Erlangen, Universität Erlangen-Nürnberg

joerg.meier@informatik.uni-erlangen.de

Kurzfassung. Zur frühzeitigen Diagnose von Erkrankungen am Augenhintergrund existieren automatische Bildverarbeitungsmethoden. Für die Effektivität solcher Verfahren und für die ärztliche Bewertung ist die Bildqualität der Eingabedaten entscheidend. Wir stellen einen neuen segmentierungsfreien Algorithmus zur automatischen Qualitätsmessung einzelner Fundusbilder vor, der Clustering-Merkmale, Schärfemaße und globale Bildstatistiken kombiniert. Die Güte der Methode wird mit einem etablierten Verfahren zur Qualitätsmessung von Retina-Fundusbildern verglichen, das nur Clustering-Merkmale auf gefilterten Bildern verwendet. Der neue Algorithmus zeigt dabei auf 302 Fundusfotos mit einer Klassifikationsgenauigkeit von 90,3% eine deutlich höhere Zuverlässigkeit als das Vergleichsverfahren mit 69,5%.

1 Einleitung

Retina-Fundusbilder stellen die Grundlage für die Diagnose von Erkrankungen am Augenhintergrund, wie Glaukom oder diabetische Retinopathie, dar. Zur frühzeitigen Erkennung existieren automatische Bildverarbeitungsmethoden, die den Arzt unterstützen sollen. Entscheidend für die Zuverlässigkeit solcher Methoden ist die Bildqualität. Um diese automatisch und sicher zu bestimmen, wird hier ein neues Verfahren vorgeschlagen. Im Gegensatz zu anderen Methoden (wie z. B. [1]) kommt die Methode ohne fehleranfällige Segmentierung von Bildstrukturen aus und bestimmt die Bildqualität eines Eingabebildes ohne ein Referenzbild zu benötigen.

Nach Wissen der Autoren gibt es drei relevante Arbeiten im Bereich von Retinabildqualitätsbewertung. Fleming et al. [1] verwenden einen segmentierungsbasierten Ansatz, bei dem zunächst die Erkennbarkeit der Gefäße im Bereich der Makula (Punkt des schärfsten Sehens) zur Messung der Qualität dient. Außerdem wird der anatomische Ausschnitt bestimmt, den das Bild zeigt. Bilder guter Qualität zeigen die Papille (Sehnervkopf) sowie den Augenhintergrund in einem Bereich der Größe des doppelten Papillendurchmessers um die Makula.

Lalonde et al. [2] berechnen die Qualität eines Eingabebildes segmentierungsfrei über das Histogramm des Gradientenmagnitudenbildes und über lokale Histogramminformationen der Grauwerte. Die Klassifikation erfolgt über Vergleiche mit einem Referenzhistogramm entwickelt aus Bildern guter Qualität.

Niemeijer et al. [3] schlagen "Image Structure Clustering" (ISC) als ebenfalls segmentierungsfreie Methode vor. Die Qualität einer Retinaaufnahme wird dabei durch die Verteilung von Clustern charakterisiert, die das Aussehen bestimmter Strukturen auf den Bildern modellieren. Hierzu wird das Eingabebild in das Gauge-Koordinatensystem transformiert, das in jedem Punkt durch die Richtung seiner Gradienten definiert ist. Nach einem k-Means-Clustering werden die Clustergrößen sowie jeweils ein 5-Bin-Histogramm jedes Farbkanals als Merkmale verwendet. Durch die Verwendung der unterschiedlich gefilterten Gauge-Koordinaten, gehen bei diesem Verfahren hauptsächlich lokale Kanteninformationen in die Merkmalsberechnung ein. Globale Bildstatistiken fehlen.

Unser Verfahren verwendet ebenfalls ein unüberwachtes Clustering, das jedoch direkt auf den Grauwerten arbeitet und so versucht, strukturelle Komponenten zu finden. Als Merkmale dienen zusätzlich zu den Clustergrößen die Inter-Cluster-Differenzen, um die Trennbarkeit der Strukturen zu messen. Außerdem führen wir ein Schärfemaß ein und berechnen drei Haralick-Merkmale [4]. Damit verbindet es lokale mit globalen Bildmerkmalen.

2 Methoden

Der Beurteilung der Bildqualität liegen vier Qualitätskriterien zu Grunde: (i) Homogene Ausleuchtung, (ii) Erkennbarkeit der Papillenstruktur, (iii) Erkennbarkeit der Gefäßstrukturen und (iv) heller bzw. kontrastreicher Hintergrund.

Der Algorithmus berechnet drei Merkmalsgruppen. Im ersten Teil wird ein unüberwachtes Clustering durchgeführt, das die Bildkomponenten in Clustern zusammenfasst. Es werden die Clustergrößen und Inter-Cluster-Differenzen bestimmt. So wird Erkennbarkeit und Differenzierung der einzelnen Strukturen bewertet. Im zweiten Teil wird ein Schärfemaß berechnet, das mit einem Schwellwertverfahren die Kantengüte schätzt. Es bewertet, wie deutlich die Strukturen abgegrenzt sind. Im dritten Teil werden durch Haralick-Merkmale die Homogenität, der Kontrast und die allgemeine Bildschärfe gemessen. Schließlich werden die drei Merkmalsgruppen zu einem Vektor kombiniert.

2.1 Clustering

Es erfolgt ein k-Means-Clustering des Eingabebildes B der Größe $n \times m$ mit k Clustern C_i , wobei $i \in \{1, \dots, k\}$. Es wird nur der Grünkanal der Bilder betrachtet, da dieser den besten Kontrast aufweist. Die Grauwerte b_{xy} , mit $x \in \{1, \dots, n\}$ und $y \in \{1, \dots, m\}$, werden zu Clustern zusammengefasst. Damit stützt sich das Verfahren nicht auf lokale Kanteninformationen. Das Clustering soll verschiedene Strukturen des Augenhintergrundes detektieren. Dazu werden die Clusterzentren durch gemittelte Werte für die k zu findenden Strukturen aus *guten* Bildern (nach Goldstandard durch Arzt festgelegt) initialisiert. Das Verhältnis c_i der Pixel b_{xy} jedes Clusters C_i zur Gesamtanzahl der Pixel wird als Kennzahl für die Größe der jeweiligen Struktur verwendet und im folgenden *Clustergröße* genannt. Als entscheidende Merkmale kommen nun die Inter-Cluster-Differenzen hinzu. Je deutlicher Strukturen erkennbar und separierbar

sind, desto höher ist der Kontrast zwischen ihnen, der durch die Inter-Cluster-Differenzen gemessen wird. Dafür wird für die Differenz aus dem Mittelwert m_i des jeweiligen Clusters C_i mit den Mittelwerten der anderen Cluster d_{ij} gebildet.

$$c_i = \frac{\#\{b_{xy} | b_{xy} \in C_i\}}{n \cdot m} \quad (1)$$

$$d_{ij} = m_i - m_j, i \in \{1, \dots, k\}, j \in \{1, \dots, k\}, i > j \quad (2)$$

Die Clustergrößen c_i und die Inter-Cluster-Differenzen d_{ij} beurteilen somit die strukturelle Erkennbarkeit der einzelnen Bildkomponenten, wie z. B. der Papille.

2.2 Schärfemessung

Zur Messung der Bildschärfe wird eine Blind-SNR-Schätzung vorgenommen. Zunächst wird die Gradientenmagnitude G des Eingabebildes B berechnet.

$$G = \sqrt{B_x^2 + B_y^2} \text{ mit } B_x = \frac{\partial B}{\partial x}, B_y = \frac{\partial B}{\partial y} \quad (3)$$

Die Grauwerte g_{xy} des Gradientenbildes G werden auf den Bereich $[0 \dots 1]$ skaliert. Zusätzlich zur Anzahl s_1 wird die Summe s_2 der Grauwerte g_{xy} berechnet, die über einem Schwellwert $\alpha \in [0 \dots 1]$ liegen. Die beiden Kennzahlen werden zur Normalisierung durch die Gesamtanzahl der Bildpunkte dividiert.

$$s_1 = \frac{\#\{g_{xy} | g_{xy} \geq \alpha\}}{n \cdot m} \text{ und } s_2 = \frac{\sum_{i=1}^{n \cdot m} v_{xy}}{n \cdot m}, v_{xy} = \begin{cases} 0 & g_{xy} < \alpha \\ g_{xy} & g_{xy} \geq \alpha \end{cases} \quad (4)$$

Der Schwellwert α wurde auf 0.5 festgelegt. s_1 beschreibt nun die Anzahl der Kantenpunkte und s_2 die durchschnittliche Kantenstärke. Beide Merkmale messen damit die Abgrenzung der einzelnen Strukturen.

2.3 Haralick-Merkmale

Es werden die drei Haralick-Merkmale [4] Entropie h_1 , Energie h_2 und Kontrast h_3 verwendet.

2.4 Kombination der Merkmale

Die Clustergrößen c_i der k Clustern, die Cluster-Abstände d_{ij} , die beiden Schärfekennzahlen s_1 , s_2 und die drei Haralick-Maße h_1 , h_2 und h_3 werden zu einem Merkmalsvektor zusammengefasst. Durch Evaluierung der Klassifikationsleistung wählten wir $k = 5$. Strukturell wird somit eine Aufteilung des Bildes in Gefäße, zwei Papillen- und zwei Hintergrundbereiche ermöglicht. Die Cluster korrelieren also gut mit den genannten Qualitätskriterien. Der resultierende 20-dimensionale Merkmalsvektor kann ohne weitere Dimensionsreduktion klassifiziert werden.

Tabelle 1. Klassifikation von *guten* Bildern: neues Verfahren und ISC

Verfahren	Klassifikationsgenauigkeit	Sensitivität	Spezifität
neues Verfahren	90,3%	95,1%	70,9%
ISC	69,5%	69,0%	70,0%

3 Ergebnisse

Zur Evaluierung diente eine Menge von 302 Retinaaufnahmen (Kowa non-myd, 22,5° FOV, linke und rechte Augen), die von drei unabhängigen Bewertern, darunter ein Augenarzt, anhand der in Abschnitt 3 genannten vier Kriterien bewertet wurden. Es erfolgte eine Einteilung in zwei Klassen *gut* und *schlecht*. Ein Bild gilt als *gut*, wenn mindestens 2 Kriterien positiv bewertet wurden. Der vorgestellte Algorithmus wurde mit dem selbst nachimplementierten ISC-Verfahren verglichen. Bei allen Tests wurde eine Support Vector Machine mit radialer Basisfunktion $k(\mathbf{u}, \mathbf{v}) = \exp(-\gamma \cdot |\mathbf{u} - \mathbf{v}|^2)$ als Klassifikator verwendet. Die Parameter γ und der Bestrafungsfaktor C wurden auf einer Teilung der Bilder in Trainings- und Testmenge optimiert. Zur Evaluierung der Klassifikation wurden je zwei Teilmengen der Bildmenge in 10 Iterationen kreuzvalidiert und daraus der Mittelwert ermittelt.

Das neue Verfahren erreicht eine Klassifikationsgenauigkeit von 90,3%. ISC erzielt 69,5% (Tab. 1). Die Sensivität des neuen Algorithmus liegt bei 95,1% und die des ISC bei 69,0%. Die Spezifität beider Verfahren unterscheidet sich mit 70,9% beim neuen Verfahren und 70,0% bei ISC nur wenig.

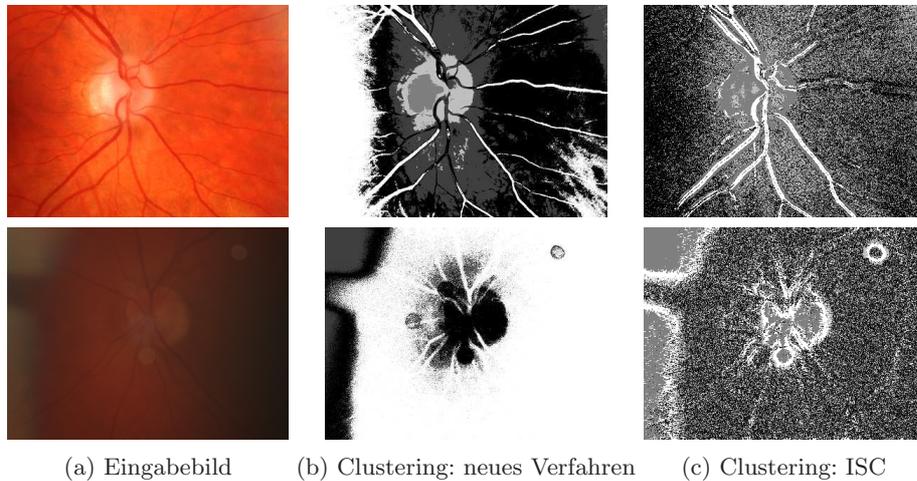


Abb. 1. Bildbeispiele: *Gute, schlechte* Retinabilder (erste, zweite Zeile) und Cluster-Ergebnis kodiert durch Grauwerte.

4 Diskussion

Trennt man die etablierten Haralick-Merkmale von den neuen Cluster- und Schärfemerkmale, erreichen die Haralick-Maße eine Sensivität von 92,7% (Spezifität von 67,0%). Die beiden neuen Cluster- und Schärfemerkmale kombiniert erzielen eine Sensivität von 74,7% (Spezifität 85,7%). Die Haralick-Merkmale alleine neigen also dazu, Bilder eher als *gut* einzustufen. Die neuen Clustering- und Schärfemerkmale dagegen identifizieren die *schlechten* Bilder besser. Kombiniert ergibt sich dadurch die Klassifikationsgenauigkeit von 90,3%. Das neue Verfahren zeigt damit eine deutlich bessere Klassifikationsgenauigkeit als das ISC. Es gibt mehrere Gründe, weshalb die publizierten Ergebnisse von ISC auf dem hier benutzten Datensatz nicht verifiziert werden konnten. In [3] wurde ISC auf Bildern mit größerem FOV als in unserem Datensatz getestet. Auch das Goldstandard Qualitätskriterium war dort anders. Trotz der niedrigeren Klassifikationsgenauigkeit, sind die Gefäße in den ISC-Clustern deutlich besser separiert (visuelle Einschätzung, siehe Abb. 1). ISC eignet sich also z. B. als Initialisierung für ein Segmentierungsverfahren besser.

Wir erreichen die hohe Klassifikationsgenauigkeit mittels Modellierung der hier geforderten Qualitätskriterien durch das neue Verfahren in den fünf Clustern und durch die Ergänzung mit globalen Bildstatistiken. Wir konnten also zeigen, dass diese Kombination von globalen und lokalen Merkmalen für die automatische Qualitätsmessung den verwendeten Goldstandard gut abbildet.

Literaturverzeichnis

1. Fleming AD, Philip S, Goatman KA, et al. Automated assessment of diabetic retinal image quality based on clarity and field definition. *Invest Ophthal Vis Sci.* 2006;47(3):1120–1125.
2. Lalonde M, Gagnony L, Boucher MC. Automatic visual quality assessment in optical fundus images. *Proc Vis Interface.* 2001; p. 259–264.
3. Niemeijer M, Abramoff MD, van Ginneken B. Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. *Med Image Anal.* 2006;10(6):888–898.
4. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern.* 1973;3(6):610–621.