

SCAN-AS-SCAN-CAN®



Eine Herausforderung beim Digitalisieren liegt darin, die gescannten Dokumente einer Datenbank oder einem ERP-System zuführen zu können. Hierzu muss der Textinhalt erschlossen werden, also das Dokument über eine Texterkennung (OCR) laufen, sodass spezifische Verschlagwortungskriterien oder Textinhalte ausgelesen werden können.

Die Art und Weise wie dies geschieht ist ausschlaggebend für die Qualität der Ergebnisse.

Optimale Texterkennung durch Farbeliminierung und ausgefeiltes Image-Processing

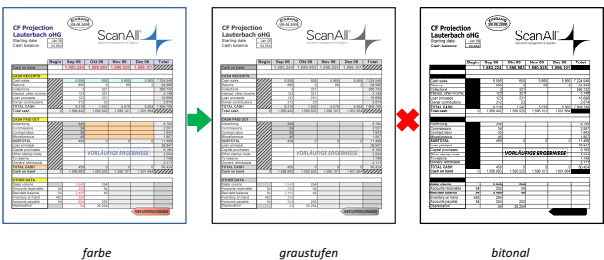
- ✦ Der Text wird vom Hintergrund separiert, also noch vor der Texterkennung freigestellt und in eine separate Ebene abgelegt. Auf diese kann dann die OCR-Maschine zugreifen.
- ✦ Dieser Textlayer wird weiter durch verschiedenste Prozesse, wie z. B. Linien- und Konturenverstärkung (besonders wichtig bei Faxen oder Nadeldrucken), speziell aufbereitet.
- ✦ Das nachfolgende Deskewing (Geraderichten) wird auf Farb- oder Graustufenebene (niemals nur in S/W, da eine 1-Bit-Information höchstens um ein Pixel verschoben, aber nicht wirklich gerade gerichtet werden kann) ausgeführt.
- ✦ Zusätzlich kommt eine Kontrastverstärkung und Gammakorrektur, sowie eine Schriftfreistellung (Umsäumung von Schriften, die z. B. durch Textmarker übermalt wurden, mit einem weißen Rand) hinzu.
- ✦ Bei Bedarf erfolgt im Image-Processing eine Eliminierung von Textmarkern oder über der Schrift liegenden Farbe.
- ✦ Farbige Schriften werden nach S/W auf 24-Bit-Basis mit entsprechenden Filtern gedithert.
- ✦ Es folgt ein verlustfreies Komprimieren des Textlayers mit neuesten und aufwändigen Wavelet-Algorithmen.

Originalgetreu gescannte Dokumente

Wird in S/W (bitonal) gescannt, so entspricht dies nicht der Qualität, wie sie vom S/W-Kopieren bekannt ist! S/W-Kopierer arbeiten mit mindestens 8-Bit, also mit einer Skala von 256 Grautönen. Ein S/W-Scan entspricht tatsächlich aber nur der Qualität eines einfachen Faxes.

Allerdings sind die zu scannenden Papierdokumente nicht nur generell wenigstens in Grauskala, sondern oft auch farbig gedruckt. Da aber beim S/W-Scannen ab einem bestimmten, der Farbe zugeordneten Grauwert (Schwellwert), dieser dann ignoriert (also als weiß angenommen) wird, gehen beim S/W-Scannen oftmals Informationen unwiederbringlich verloren. Farbiger oder grauer Text wird deswegen auch bei der nachfolgenden Texterkennung oftmals gar nicht oder fehlerhaft erkannt.

Außerdem entspricht der S/W-Scan (bitonal) eines ehemals farbigen oder in Graustufe gedruckten Papierdokuments durchaus nicht mehr dem originalen, ursprünglichen Aussehen dieses Dokuments. Bei rechtsrelevanten Dokumenten kann so die Beweiskraft eines Scans vollständig verloren gehen.



- ✦ Wir scannen deshalb nie nur in S/W (1-Bit), sondern mindestens in 8-Bit-Grauskala bzw. meistens in 24-Bit-Vollfarbe.
- ✦ Soll das Ausgabeformat S/W sein, so wird erst nach dem Image-Processing nach S/W gedithert – und dies mit den besten und aufwändigsten Filtern, um Bildstörungen durch die Quantisierung nahezu ausschließen zu können.
- ✦ Werden die Scans in Vollfarbe abgespeichert, so bleiben alle Farbinformationen erhalten – d. h. es ist dann auch im Scan z. B. noch zu erkennen, in welcher Kugelschreiberfarbe ein Dokument unterschrieben wurde, oder welche Textpassagen mit einem Textmarker und in welcher Farbe hervorgehoben waren. Trotzdem ist der Textlayer zusätzlich in einer zweiten Ebene (wie anfangs beschrieben) für eine optimierte Texterkennung (OCR) in S/W (also von der Farbinformation separiert) verfügbar.

Volltextsuche gleich inklusive

- ✦ Bei uns wird generell immer der gesamte Text eines Dokuments ausgelesen und nicht nur ein bestimmtes Feld, welches z. B. für die Indexierung relevant ist.
- ✦ Die OCR-Ergebnisse werden dann gleich mit in das PDF abgespeichert und stehen für eine (auch dokumentenübergreifende) Suche im Volltext sowie für eine eventuelle Weiterverwendung in Textbearbeitungsprogrammen (via Copy & Paste) uneingeschränkt zur Verfügung. Der Text kann sogar optional als HTML, RTF oder TXT-Datei mit ausgegeben werden.

Das alles bei kleinster Dateigröße

- ✦ Die ScanAll® PDF/A-Dateien in Vollfarbe sind nicht viel größer als ein S/W-Scan im TIF-Format.
- ✦ Auch Unternehmen, die nur S/W-Belege haben, profitieren von den ScanAll® PDF/A-Dateien! Unsere S/W-PDF/A sind um 50 % kleiner als herkömmliche TIFF-Scans!



ScanAll®-Scans entsprechen in Aussehen und Qualität also vollkommen dem ursprünglichen Papierdokument! Somit sind die von ScanAll® optimierten Scans den reinen S/W-Scans weit überlegen. Dies wirkt sich äußerst positiv auf die OCR-Ergebnisse und somit letztlich auch auf die richtige Ablage der Dokumente in der Datenbank sowie die Geschwindigkeit und Zuverlässigkeit mit der die Dokumente nach Textinhalt durchsucht werden können aus.

ScanAll® - Der Spezialist für Ihr Dokumentenmanagement!

ScanAll® ist ein führender mittelständischer Spezialist für analoges wie digitales Dokumentenmanagement, Scandienstleistungen (Imaging) und PDF-Technologie, Content-Management sowie Dokumentenlogistik.

Unser erfahrenes, engagiertes Team konzeptioniert und realisiert bereits seit über 20 Jahren innovative und leistungsstarke Dokumentenmanagement- und Digitalisierungsprojekte für Technologiegroßunternehmen wie u. a. die **Siemens AG, Infineon Technologies AG, Qimonda AG, Nokia Siemens Networks, Continental Automotive GmbH** und die **Bosch und Siemens Hausgeräte GmbH**.

Unsere Kunden setzen auf unsere qualifizierten Dienstleistungen, um hochkomprimierte, elektronische Dokumente zu erhalten.

Aufgrund der Art unserer Dienstleistungen – der Umgang mit vertraulichen und betrieblichen Dokumenten der Kunden – ist Datenschutz und Datensicherheit eines der wichtigsten Qualitätsziele unseres Unternehmens.

IT-zertifiziert nach ISO-27001 (Zert.-Nr. SOE-0031-2006)

Die ScanAll® GmbH führt das „Prima-Klima-Siegel“ des Vereins PRIMAKLIMA-WELTWEIT e. V. - als Beleg für nachhaltiges umweltgerechtes Handeln.



ScanAll® GmbH
Postfach 22 13 03
D-80503 München

Fon +49 [0] 89.66 07 78 - 01

Fax +49 [0] 89.66 07 78 - 05

mail@scanall.eu

www.scanall.eu

ScanAll®, das ScanAll®-Logo, der blaue Stern, BlueStar® und Scan-As-Scan-Can® sind rechtlich als Wort und/oder Bildmarken geschützt. Alle anderen genannten Marken sind im Besitz der jeweiligen Unternehmen oder Personen.