


Visuelle Suche in historischen Werken

Markus Brantl¹ · Klaus Ceynowa¹ · Thomas Meiers² · Thomas Wolf¹ Eingegangen: 13. Oktober 2016 / Angenommen: 9. Januar 2017 / Online publiziert: 2. Februar 2017
© Der/die Autor(en) 2017. Dieser Artikel ist eine Open-Access-Publikation.

Zusammenfassung Die Bayerische Staatsbibliothek (BSB) zählt mit ihrem Bestand von knapp 11 Mio. Bänden zu den bedeutendsten Universallibliotheken der Welt. Bereits 1,2 Mio. Werke sind digitalisiert, was die BSB zur größten digitalen Kulturinstitution in Deutschland macht. Dieser digitale Bestand umfasst vorwiegend urheberrechtsfreie Werke vom 8. bis ins 20. Jahrhundert, von der mittelalterlichen Bibelhandschrift bis zur Boulevardzeitung der 1920er-Jahre. Diese Vielfalt des zu digitalisierenden schriftlichen Kulturerbes und das hohe Tempo der Massendigitalisierung in den letzten Jahren haben ihren Preis – die inhaltliche Erschließung der Werke hinkt hinterher, insbesondere bei Werken, die nicht mittels Optical Character Recognition-Verfahren (OCR) automatisiert maschinenlesbar transformiert und zugänglich gemacht werden können. Dies gilt insbesondere für mittelalterliche Handschriften, Alte Druck- und Spezialbestände. Deshalb blieb auch der reichhaltige, in diesen Werken verborgene Bildbestand für den Nutzer weitestgehend verborgen und konnte lediglich

durch das Durchblättern am Bildschirm entdeckt werden. Dies war Motivation für die Bayerische Staatsbibliothek, gemeinsam mit dem Fraunhofer Heinrich-Hertz-Institut in Berlin ein System zur ähnlichkeitsbasierten Bildsuche aufzubauen, welches sämtliche Bildinhalte aller 1,2 Mio. Digitalisate automatisch identifiziert. Hierbei werden mittels morphologischer Verfahren Bilder aus den Buchseiten extrahiert, die danach aufgrund von Farb- und Kantenmerkmalen klassifiziert werden. Bilder „ohne Informationswert“ werden mit Hilfe von Methoden aus dem Bereich des maschinellen Lernens herausgefiltert. Damit konnten aus den digitalisierten Werken der BSB bislang mehr als 43 Mio. einzelne Bilder identifiziert werden, die mittels einer hochperformanten Suchmaschine über eine frei verfügbare Web-Applikation dem Anwender direkt zur Verfügung stehen. Dank der Vielfalt und Reichhaltigkeit der indexierten Bestände spricht dieses Angebot nicht nur Historiker und Buchwissenschaftler an, sondern Interessierte aus den unterschiedlichsten Fachrichtungen. Die Ähnlichkeitssuche stellt dabei unbekannte, ungewöhnliche und oftmals überraschende Bezüge zwischen unterschiedlichsten Werken her.

Markus Brantl
markus.brantl@bsb-muenchen.de

Klaus Ceynowa
klaus.ceynowa@bsb-muenchen.de

Thomas Meiers
thomas.meiers@hhi.fraunhofer.de

✉ Thomas Wolf
thomas.wolf@bsb-muenchen.de

¹ Bayerische Staatsbibliothek,
Ludwigstraße 16, 80539 München,
Deutschland

² Heinrich-Hertz-Institut, Fraunhofer Institut für
Nachrichtentechnik, Einsteinufer 37, 10587 Berlin,
Deutschland

1 Ausgangslage

Bereits seit 1997 werden im Münchener Digitalisierungszentrum (MDZ) der Bayerischen Staatsbibliothek (BSB) ihre reichhaltigen und weltweit einzigartigen Sammlungen in ihrer Funktion als internationale Forschungsbibliothek digitalisiert und online bereitgestellt. Die Digitalisierungsaktivitäten wurden in den letzten zehn Jahren massiv durch die Kooperation mit Google Books ausgeweitet¹. Die Zahl

¹ <https://www.bsb-muenchen.de/ueber-uns/kooperationen/google/>.

der online verfügbaren, urheberrechtsfreien Werke ist auf nunmehr 1,2 Mio. angewachsen. Auf der Zeitschiene handelt es sich dabei um Bibliotheksbestände vom 8. bis ins 20. Jahrhundert, also von der mittelalterlichen Bibelhandschrift bis zur Boulevardzeitung der 1920er-Jahre. Diese Vielfalt des zu digitalisierenden schriftlichen Kulturerbes und das hohe Tempo der Massendigitalisierung in den letzten Jahren hat ihren Preis – die inhaltliche Erschließung der Werke hinkt hinterher, insbesondere bei Werken, die nicht mittels Optical Character Recognition-Verfahren (OCR) automatisiert maschinenlesbar transformiert und zugänglich gemacht werden können. Dies gilt insbesondere für mittelalterliche Handschriften, Alte Druck- und Spezialbestände. Zwar sind alle Werke der BSB – vom Handschriftenfragment des 7. Jahrhunderts bis in die Moderne – durch umfangreiche, online durchsuchbare *bibliographische* Metadaten indexiert und recherchierbar, aber die Sichtung und Aufarbeitung der Struktur eines digitalisierten Buches, also die Erfassung von Überschriften, Kapitelnummern, Illustrationen und Beilagen, ist im Fall der oben genannten Bibliotheksbestände eine aufwendige und teure Handarbeit. Somit bleibt in der Regel der bildliche Inhalt eines Buches dem Nutzer solange verschlossen, bis er das Werk am Bildschirm Seite für Seite durchblättert.

Ein weiterer innovativer Aspekt, der zu dem Projekt Bildsuche führte, war das Experimentieren mit einem alternativen, nicht-textuellen Zugang zum schriftlichen kulturellen Erbe. Dies führte im Jahr 2011 zu einer Marktsichtung durch das MDZ, in deren Folge gemeinsam mit dem Fraunhofer Heinrich-Hertz-Institut (HHI) in Berlin das innovative und seinerzeit weltweit einmalige Projekt begonnen werden konnte².

Mit dem Fraunhofer Heinrich-Hertz-Institut hat die BSB einen kompetenten Partner gefunden, der Dank langjähriger Erfahrung und seines Know-How mit Bilderkennungs- und Bildanalyse-Systemen ebenfalls bereit war, neue Wege in der Softwareentwicklung für den Kulturbereich zu gehen. So konnte noch 2011 eine erste prototypische Applikation zur Bildähnlichkeitssuche für 250 digitalisierte Bücher der BSB umgesetzt werden. Dieser erste Anwendungsfall zeigte die noch bestehenden Herausforderungen auf. Denn die Bilderkennungs-Software – bislang ausschließlich für die Plagiat-Detektion moderner Digitalfotografien eingesetzt – musste in einem aufwändigen gemeinsamen Prozess an die Besonderheiten der digitalisierten Bestände der BSB angepasst werden. Es galt, einer großen Bandbreite von heterogenen Inhalten und den Besonderheiten der digitalen Bildproduktion bei der Retrodigitalisierung gerecht zu werden. Von kunstvollen Miniaturen mittelalterlicher Handschriften über prachtvoll kolorierte Holzschnitte der Neuzeit bis hin zu feinen Stahlstichen und grob gerasterten Fotografien in

Büchern des 19. und frühen 20. Jahrhunderts ist jedes denkbare Material dabei. Zusätzlich stellte sich die Frage: Wie geht man mit dem „störenden“ Text um, der zunächst jede Bildähnlichkeitssuche verfälschte?

Nach vielen Monaten intensiver Entwicklungsarbeit konnte im April 2013 schließlich das erste Angebot der BSB zur Bildähnlichkeitssuche online gehen. Bereits damals standen rund 4 Mio. einzelne Bildsegmente aus 60.000 Büchern für eine Ähnlichkeitssuche zur Verfügung. Diese Zahl konnte in den folgenden Jahren auf 6 Mio. Bildsegmente aus 80.000 Bänden gesteigert werden. Im Jahr 2016 wurde eine neue Phase eingeläutet. Die neueste Version der Software erfasst alle Digitalisate der BSB und bietet nun 43 Mio. Bilder zur Suche an.

2 Technische Grundlagen der Bildsuche

Buchseiten können eine Vielzahl von Abbildungen enthalten, so z. B. Illuminationen, Grafiken, Fotografien, Zeichnungen und Schmuckornamente. Ziel ist es, diese Bilder über alle Buchbestände hinweg durchsuchbar zu machen. Die Grundlage der Bildähnlichkeitssuche bildet die visuelle Ähnlichkeit von Bildern, das heißt Bilder werden aufgrund ihrer Farb-, Kanten- und Texturmerkmale miteinander verglichen. Als Suchverfahren wird „Query by example“ angewandt, wobei ein Suchbild vorgegeben wird, zu dem die visuell ähnlichsten Bilder im Bestand recherchiert werden. Die visuelle Suche wird oft als iterativer Vorgang genutzt, bei dem sich der Nutzer schrittweise an die gesuchten Bilder herantastet. Die Ähnlichkeitssuche wurde als Client-Server-System implementiert. Der Nutzer startet die Bildsuche über einen Webclient. Die Anfrage wird von einem Server bearbeitet, der das Ergebnis an den Webclient zurückgibt.

Um Suchvorgänge zu ermöglichen muss der Bestand der digitalisierten Bücher entsprechend aufbereitet bzw. indexiert werden. Die Indexierung, die auf dem Server durchgeführt wird, umfasst die folgenden Schritte:

- Text-Bild-Trennung.
- Extraktion von visuellen Deskriptoren.
- Ausfiltern von „nicht relevanten Bildern“.

Wird der Buchbestand mit einer Bildsuche versehen, muss zunächst das gesamte Material indexiert werden.

Die Aufbereitung des Materials erfolgt in folgenden Schritten:

2.1 Text-Bild-Trennung

Bei der Text-Bild-Trennung wird auf Verfahren aufgesetzt, die Textflächen von Nichttextflächen auf Buchseiten trennen. Diese Verfahren dienen oftmals als Vorstufe einer

² <https://bildsuche.digitale-sammlungen.de>.

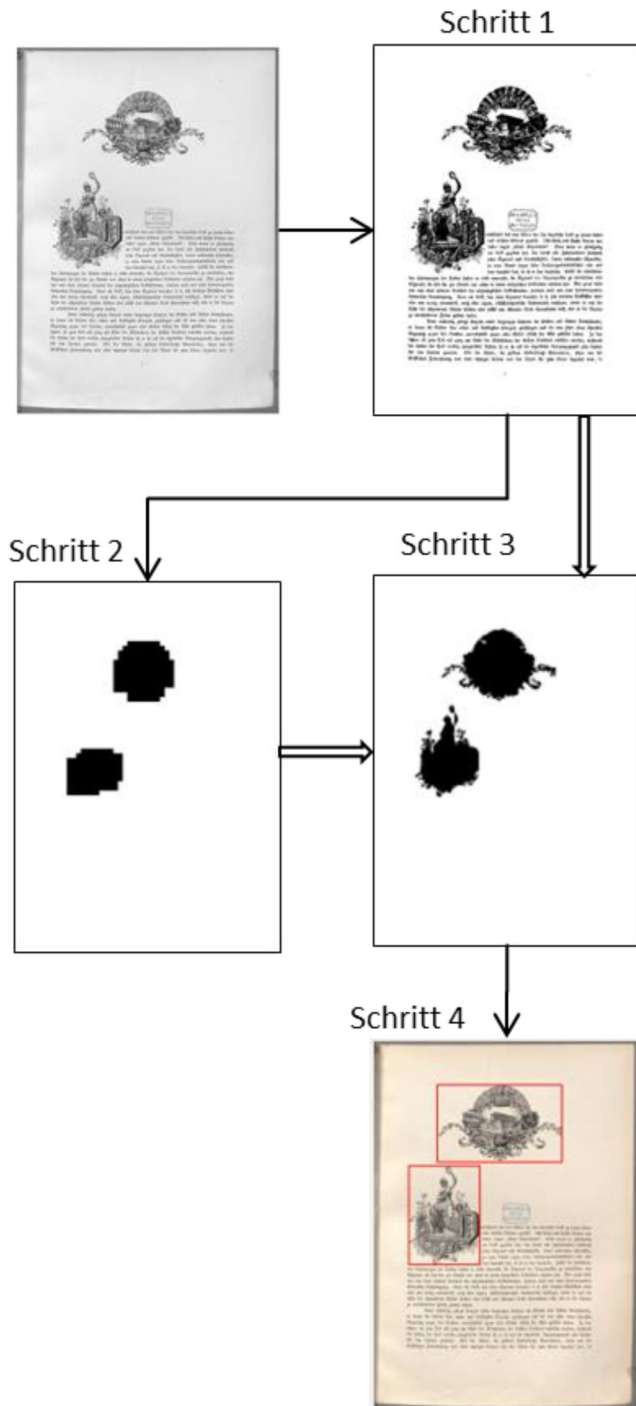


Abb. 1 Verarbeitungsschritte bei der Text-Bildtrennung

OCR-Segmentierung. Hier werden sie genutzt, um Bilder auf Buchseiten zu extrahieren. Es wird ein Ansatz verwendet, der auf einer Folge morphologischer Operationen basiert. Das Verfahren selbst ist an die Arbeiten [1, 2] angelehnt. Es wird im Folgenden grob die Vorgehensweise beschrieben.

Im ersten Schritt wird eine Otsu-Binarisierung des Grauwertbildes durchgeführt (siehe Abb. 1, Schritt 1). Liegt der

Helligkeitswert eines Pixels unter einem Schwellwert, wird der Wert auf 1 (Vordergrund), sonst auf 0 (Hintergrund) gesetzt.

Es werden hauptsächlich morphologische Operationen verwendet, welche einen 2×2 Pixel umfassenden Block durch ein einzelnes Pixel ersetzen. Ist die Summe der Pixelhelligkeitswerte des Blocks größer oder gleich einem Schwellwert T , der zwischen 1 und 4 liegen kann, so erhält der neue Pixel den Wert 1, sonst 0. Durch diese morphologische Operation wird das Bild um den Faktor 2 in Breite und Höhe verkleinert, wobei je nach Größe des Schwellwertes kleine Strukturen verschwinden. Diese Operation wird viermal hintereinander mit den Schwellwerten 1, 1, 4 und 3 durchgeführt, wodurch das Bild verkleinert wird und „kleine“ Strukturen wie Buchstaben verschwinden (siehe Abb. 1, Schritt 2).

Übrig bleiben Bereiche, die größere zusammenhängende Strukturen wie Linien und Flächen beinhalten. Diese Bereiche werden mit dem originalen Binärbild überlagert. Mit der morphologischen Operation „Dilatation“ werden Lücken gefüllt (siehe Abb. 1, Schritt 3). Als Ergebnis erhält man zusammenhängende Bereiche, die jeweils zu einem Rechteck aufgefüllt werden (siehe Abb. 1; Schritt 4 Rahmen). Diese Rechtecke umfassen die Nicht-Textflächen und kommen als Bilder oder Grafiken in Frage.

2.2 Extraktion visueller Merkmale

Die Suche erfolgt nicht über die Bilder selbst, da dies bei einem Bestand von circa 43 Mio. nicht in Echtzeit durchführbar wäre; vielmehr kommen Deskriptoren zum Einsatz. Deskriptoren sind Datensätze, die in sehr komprimierter Form die visuelle Information eines Bildes enthalten. In unserem Fall hat der zu einem Bild gehörende Deskriptor einen Umfang von 96 Byte. Zusätzlich wird eine Distanzfunktion benötigt, die den Abstand zweier Deskriptoren angibt. Diese Funktion soll über die Deskriptoren den visuellen Unterschied zweier Bilder möglichst optimal abbilden. Aus der Distanzfunktion wird die Ähnlichkeitsfunktion berechnet, die einen Wert zwischen 0,0 und 1,0 ausgibt. Der Wert 0,0 bedeutet dabei maximale Unähnlichkeit, der Wert 1,0 maximale Ähnlichkeit bzw. Identität.

Die Erforschung geeigneter globaler Deskriptoren ist Gegenstand vieler Arbeiten; einen Überblick ist in der Arbeit [3] dargestellt.

Die Aufstellung von Deskriptoren zur Beschreibung von Medieninhalten war auch Gegenstand der MPEG-7 Standardisierung, die im Jahre 2001 verabschiedet wurde. Jedoch wurde in diesem Standard für jedes visuelle Merkmal (Farbe, Farbverteilung, Textur, Kantenverteilung etc.) ein eigener Deskriptor entwickelt. Unser Ziel war es einen eigenen Deskriptor zu schaffen, der mehrere visuelle Merkmale

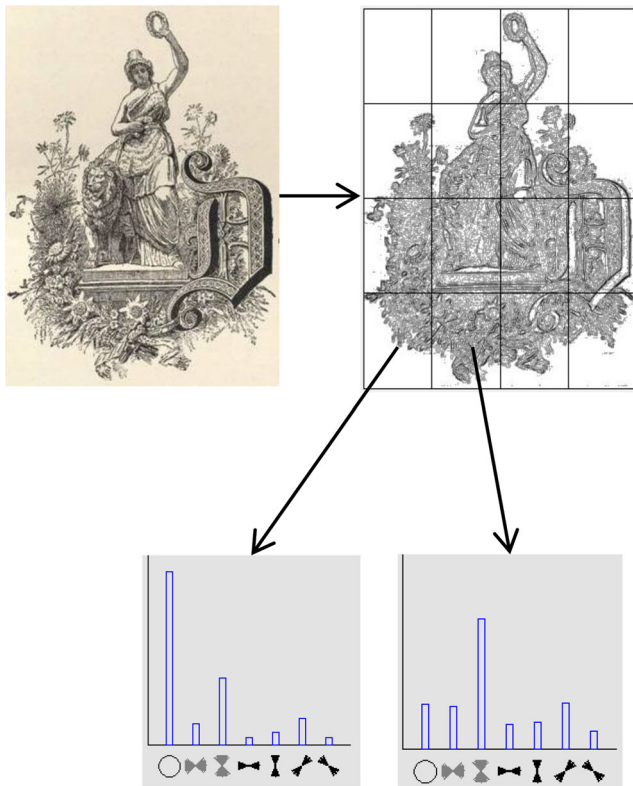


Abb. 2 Kanten/Texturhistogramme für die einzelnen Bereiche. Oben links ist das Originalbild, rechts das Kantenbild, darunter Kanten/Texturhistogramme für zwei Bildbereiche. Die Abszisse gibt die verschiedenen Orientierungen an

gleichzeitig abdeckt und performanter sowohl bei der Extraktion als auch bzgl. der schnellen Suche ist.

Der visuelle Deskriptor enthält Informationen bezüglich der Farb- als auch der Verteilung der Kanten/Texturorientierung des Bildes. Zur Erfassung der Farbinformation wird das Bild in 8×8 einheitliche Bereiche zerlegt. Für jeden Bereich werden der durchschnittliche Grauwert (Y-Wert) sowie die Farbwerte Cb und Cr ermittelt. Auf diese Weise erhält man 3 Blöcke mit 8×8 Werten, die jeweils einer zweidimensionalen Kosinustransformation unterworfen werden. Die so erhaltenen Koeffizienten werden anhand der Frequenz sortiert und die ersten 15 (Grauwert Y) bzw. 10 Koeffizienten (Farbwerte Cb und Cr) werden als Wert für den Deskriptor übernommen. Die Farbinformation wird durch 35 Werte, jeweils über 8-bit (1 Byte) quantisiert, dargestellt.

Zur Darstellung der Kanteninformation wird zunächst ein Kantenfilter auf das Grauwertbild angewandt. Man erhält für jeden Pixel einen Kantenvektor. Ist keine Kante vorhanden, so ist der Kantenvektor ein Nullvektor. Die Kanten werden je nach Länge in drei Klassen eingeteilt:

- Konturlose Flächen (Kantenvektoren sehr klein, in der Abszisse des Histogramms in Abb. 2 als Kreis dargestellt).
- Texturen (Kantenvektoren haben einen mittleren Wert, in Abb. 2 graues Symbol in der Abszisse).
- Echte Kanten (Kantenvektoren haben einen hohen Wert, in Abb. 2 schwarzes Symbol in der Abszisse).

Bei Texturen werden zwei (senkrecht und waagrecht), bei den echten Kanten vier Richtungen unterschieden (senkrecht, waagrecht sowie die beiden diagonalen Richtungen). Das Bild wird in 16 Bereiche unterteilt (jeweils vier in senkrechter wie in waagerechter Richtung). Für jeden Bereich wird die Häufigkeit der verschiedenen Kanten (Klassen wie Richtungen) ermittelt. Man erhält 16 Kanten/Texturhistogramme mit jeweils 7 Werten, also insgesamt 112 Werte, die jeweils in einen 4-bit langen Wert quantisiert werden. Die Berechnung der Kanten/Texturinformation als auch die Distanzfunktion sind durch zwei Patentfamilien geschützt (s. deutsche Patente [4]).

Die Werte aus dem Farblayout sowie die Häufigkeiten der Kanten aus allen 16 Bereichen werden zu einem 147-dimensionalen (35 für die Farb- und 112 für die Kanten/Texturinformation) Vektor zusammengefasst. Als Distanzmaß wurde die L2-Norm für die einzelnen Farbraumkomponenten Y, Cb und Cr sowie für die Kanten/Texturwerte getrennt berechnet und gewichtet aufaddiert. Bei der Distanzberechnung kann die Gewichtung zwischen den Farb- und den Kanten/Texturwerten vorgegeben werden. Diese Norm hat sich in Tests gegenüber der L1 bzw. Maximumnorm als geeigneter erwiesen. Über eine $\exp(-x)$ -Funktion wird dieses Maß auf eine Ähnlichkeitsskala von 0,0 bis 1,0 abgebildet..

Die Länge des Deskriptors beträgt 96 Byte. Der geringe Umfang des Deskriptors hat zur Folge, dass nur ein geringer Speicherplatz benötigt wird und die Distanzberechnungen sehr schnell erfolgen können (mehrere Millionen Distanzberechnungen pro CPU-Kern in der Sekunde).

2.3 Aussortieren von nicht relevanten Bildern mit Verfahren des maschinellen Lernens

Die bisherige Nutzung hat gezeigt, dass ein Teil der extrahierten Nicht-Textstellen und Bilder für den Anwender ohne Informationswert ist. In einer nachgelagerten Analyse werden diese „Bilder ohne Informationswert“ mit Hilfe von Verfahren aus dem Bereich des maschinellen Lernens herausgefiltert.

Hierzu wurden zunächst die folgenden 11 Klassen von Bildern ohne Informationswert zusammengestellt:

Einfarbige Flächen, Buchdeckel, leere Seiten, Flecken, zwei Varianten von Farbskalen, Klammern, Tabellen, Musiknoten und zwei Varianten von Textstellen. Anschließend

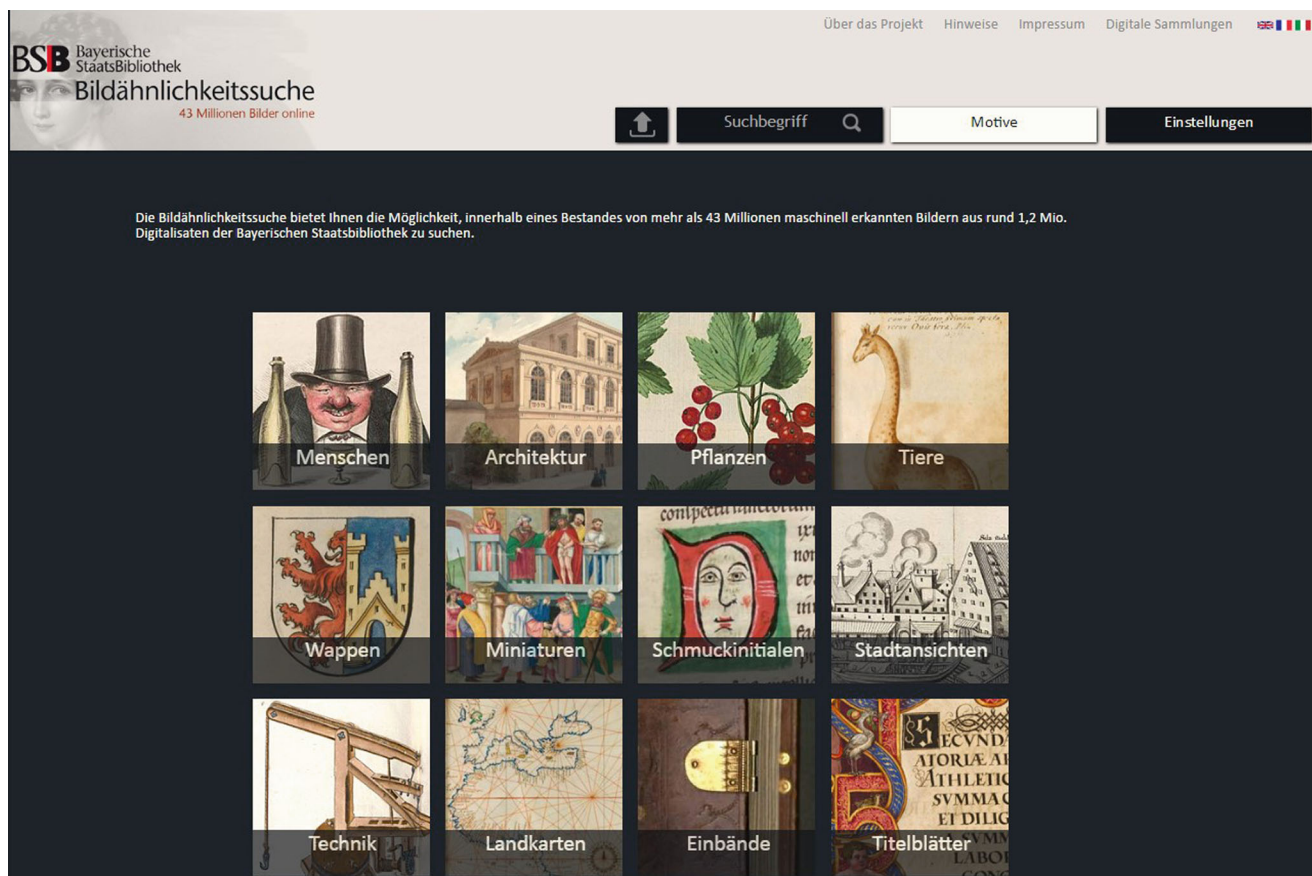


Abb. 3 Frontend-Applikation der Bayerischen Staatsbibliothek

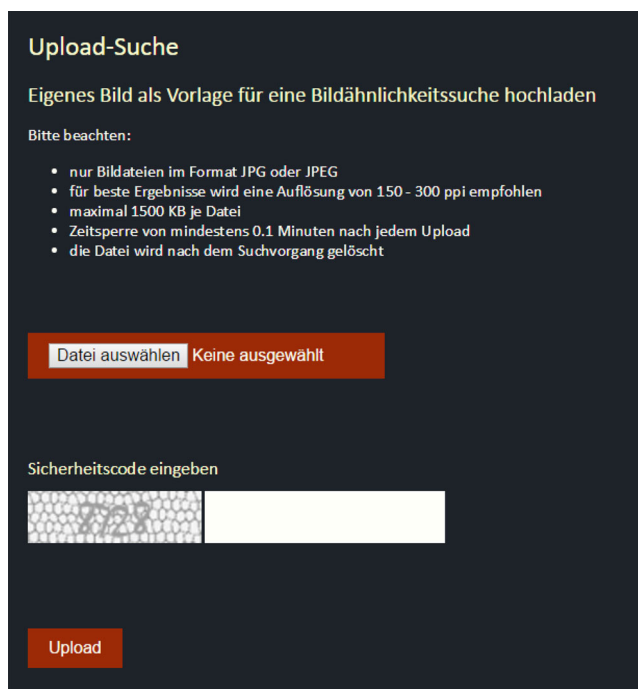


Abb. 4 Upload-Modul der Bildähnlichkeitssuche

wurde für jede Klasse eine einklassige Support-Vector-Maschine (SVM) mit einer Gauß-Funktion als Kernel mit den Deskriptoren der Beispielbilder trainiert.

Das System ist jederzeit erweiter- und anpassbar an neue Klassen. Zur Beurteilung, ob ein Bild behalten oder aussortiert werden soll, wird der Deskriptor nacheinander in alle SVMs eingespeist. Erkennt eine SVM die Zugehörigkeit eines Bildes zu ihrer Klasse, wird es aussortiert. Die Klassifizierung erfolgt sehr schnell, da die Anzahl der Supportvektoren pro SVM im zweistelligen Bereich liegt. Bei 11 SVMs sind Distanzberechnungen in der Größenordnung von circa tausend Berechnungen notwendig, die eine Rechenzeit von ca. einer Millisekunde bei Nutzung eines Kernes eines üblichen PC-Prozessors benötigen.

2.4 Effiziente parallele Suche

Die Deskriptoren werden in einer Indexdatei auf dem Server abgespeichert. Neben den Deskriptoren werden aber auch Informationen zu dem dazugehörigen Bild, der Buchseite und dem Buch, aus welchem das Bild stammt, benötigt. Neben diesen Informationen muss noch die Position auf der Buchseite in Pixelkoordinaten angegeben werden. All diese Informationen werden in einer Indexdatei angegeben.

Abb. 5 Beispiel für Suchvorlage (*links*) und Treffer



Beim Start der Anwendung wird die Indexdatei in den Arbeitsspeicher geladen. Liegt eine Suchanfrage vor, wird der Deskriptor ermittelt. Ist das Bild der Suchanfrage aus der Menge des Bestandes, kann der Deskriptor aus den Indexdaten direkt eruiert werden. Der Deskriptor der Suchanfrage wird mit den Deskriptoren des Bestandes verglichen. Die jeweils k besten Ergebnisse werden ausgegeben. Die Anzahl k der besten Ergebnisse kann vom Nutzer eingestellt werden.

Für Suchvorgänge können Indexstrukturen verwendet werden. So wurde der M-tree, der auf beliebige Metriken anwendbar ist, auf seine Eignung geprüft. Beim M-tree werden die Deskriptoren in einem hierarchischen Baum angeordnet, der bei der Suche durchlaufen wird. Aufgrund der hohen Dimensionalität war der Vorteil für die Suche nicht so hoch wie erwartet. Auf der anderen Seite wuchs der Aufwand für den Aufbau der M-trees unverhältnismäßig stark. Auch ist die Parallelisierung von Baumstrukturen aufwändig. Die lineare Suche zeigte sich aufgrund der hohen Dimensionalität und Datenmenge als die praktikablere Lösung.

Da zurzeit 43 Mio. Bilder und damit die gleiche Anzahl Deskriptoren vorliegen, müssen pro Suchanfrage 43 Mio. Deskriptorenvergleiche durchgeführt werden. Um die Suchanfrage zu beschleunigen, kann sie parallel über beliebig viele CPU-Kerne durchgeführt werden. Hierzu wird die Liste der Bestands-Deskriptoren in gleich große Teillisten aufgeteilt, die jeweils auf einem CPU-Kern durchlaufen werden. Danach werden die Einzelergebnisse zu einem Ergebnis fusioniert. Es hat sich gezeigt, dass eine Aufteilung einer Suchanfrage auf vier CPU-Kerne ausreichend ist und zu reinen Suchzeiten unter 1 s führt.

3 Die Applikation

Für den Endbenutzer wurde seitens der Bayerischen Staatsbibliothek eine einfach zu bedienende Web-Applikation (siehe Abb. 3) in responsivem Design programmiert, so dass sie auf unterschiedlichen Endgeräten verwendet werden kann.

Diese Applikation wird pro Woche von durchschnittlich 3000 Besuchern aus aller Welt genutzt, welche zusammen mehr als 50.000 Seitenaufrufe tätigen.

Jede Ähnlichkeitsrecherche über den vorhandenen Bestand benötigt ein Ausgangsbild als Suchvorlage.

Dies kann einerseits über die zentrale Einstiegsseite <http://bildsuche.digitale-sammlungen.de> geschehen oder über eine Buchseite, die ein User im Standard-Bildviewer des MDZ betrachtet.

Für den zentralen Sucheinstieg, der den Zugang für alle 43 Mio. Bilder bietet, stellte sich die Frage, wie man den bildbasierten Einstieg für den User sinnvoll gestalten kann. Denn es ist verschiedenen Nutzungsszenarien Rechnung zu tragen. Manche Benutzer suchen gezielt nach einem bestimmten Motiv, andere möchten lediglich stöbern, um die Bestände erst einmal kennenzulernen.

3.1 Einstiege

Die erste und meistgenutzte Option des Einstiegs ist ein Pool aus einigen tausend charakteristischen Bildvorlagen, die zusätzlich nach Kategorien geordnet sind und sich als Einladung zum Stöbern und Entdecken verstehen. Hier findet man zum Beispiel eine Auswahl von Porträts aus unterschiedlichen Epochen, außerdem Architekturzeichnungen und Abbildungen von Bauwerken, Illustrationen von Pflanzen und Tieren, sowie Landkarten, Wappen, Bucheinbände, Titelblätter und Exlibris. Schließlich gibt es noch die Option, sich Bilder als reine Zufallsauswahl anzeigen zu lassen.

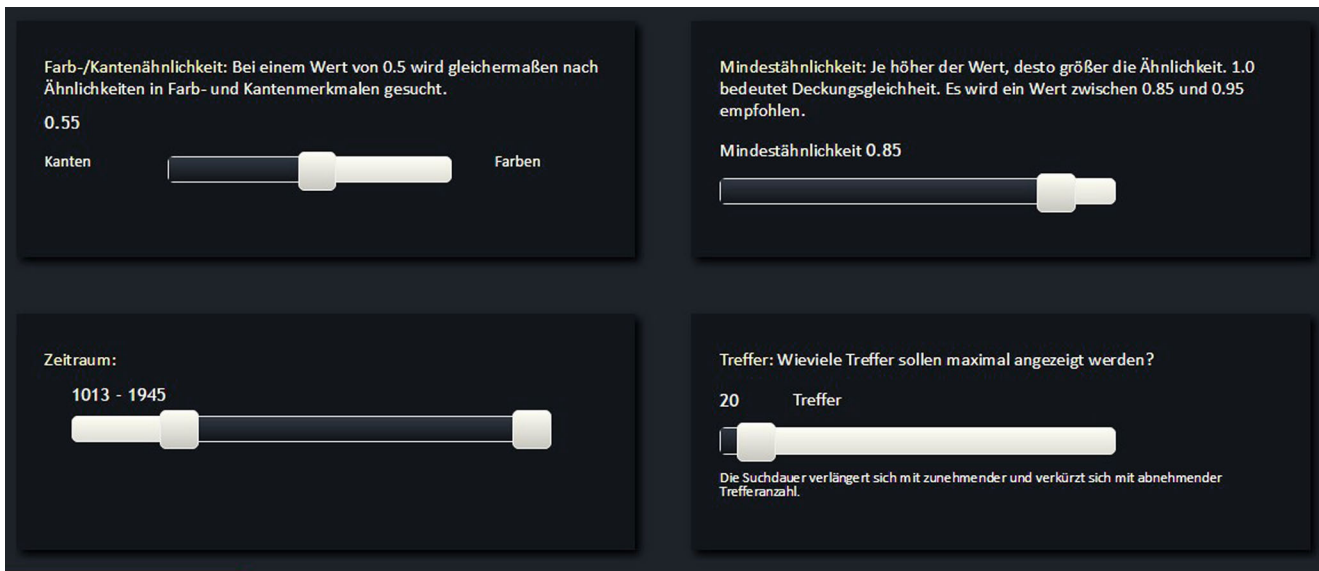


Abb. 6 Optionen zur Manipulation der Suche

Bei dieser Form des Einstiegs vollzieht sich die eigentliche Ähnlichkeitssuche erst ab dem zweiten Schritt, wenn ein bestimmtes Motiv ausgewählt wurde, welches zu ähnlichen Motiven führt, die wiederum als Suchvorlage verwendet werden.

Ein alternativer Einstieg in die Bildähnlichkeitssuche wurde zusätzlich außerhalb der eigentlichen Applikation verankert, nämlich im allgemeinen Viewer, der den Nutzern zum Durchblättern und Betrachten aller Digitalisate der BSB dient. Über eine Schnittstelle im Hintergrund wird bei jedem Blättervorgang geprüft, ob für die aktuell geöffneten Seiten Informationen im Bildindex abgelegt sind. Ist dies der Fall, kann unmittelbar eine Ähnlichkeitssuche mit Inhalten dieser Seite gestartet werden.

3.2 Upload – Eigene Bilder mit dem BSB-Bildindex abgleichen

Die vielleicht wichtigste Funktion für den Nutzer besteht in der Möglichkeit, eine beliebige eigene Bildvorlage per Upload als Suchvorlage einzusetzen (siehe Abb. 4). Somit ist der Endbenutzer nicht darauf angewiesen, bei der Suchvorlage nur *innerhalb* des Bildbestandes der BSB fündig zu werden. Man kann stattdessen Abbildungen aus dem eigenen Kontext der Forschungsarbeit mit dem Bestand der BSB abgleichen.

Der Heraldiker verwendet seine eigene Bildersammlung, um auf diese Weise im reichhaltigen Bestand an Wappenbüchern der BSB fündig zu werden. Der Buchwissenschaftler entdeckt Initialen und Abbildungen oder gleicht unterschiedliche Titelblätter aus Druckwerken des 16. Jahrhunderts miteinander ab. Der Zeitungswissenschaftler findet Werbeanzeigen für Haushaltswaren und Kraftfutter. Die

von den Nutzern hochgeladenen Bilder werden dabei nur temporär verwendet und nach der Suche umgehend wieder gelöscht. Die Upload-Option gehört zu den beliebtesten Funktionen des Online-Angebotes.

Welche Möglichkeit des Einstiegs der Nutzer auch wählt – fortan kann jeder einzelne Suchtreffer für sich als Ausgangsbild einer neuen Ähnlichkeitssuche verwendet werden (siehe Abb. 5). Auf diese Weise dringt man immer weiter in den Bildbestand vor und macht nützliche, überraschende und teils auch skurrile Entdeckungen. Mittelalterliche Miniaturen ähneln Zeitungsfotos. Dieses sich Vertiefen und sich Verlieren ist dabei intendiert. Gleichwohl stellt das Angebot eine gewisse Herausforderung dar und die Geduld des Nutzers häufig auf die Probe. Denn nicht jeder Suchtreffer ergibt semantisch einen Sinn, auch wenn der Algorithmus der Anwendung plausible Ähnlichkeiten entdeckt.

Der Nutzer kann bei der Suche im Frontend den Einfluss der Farb- gegenüber den Kanten- und Texturmerkmalen gewichten (siehe Abb. 6). In der Regel ergeben sich bei einer Suche nach Kantenähnlichkeit oder bei einem ausgewogenen Verhältnis von Farb- und Kantenmerkmalen die besten Treffer. Dagegen ergibt eine Suche nach reiner Farbähnlichkeit in den meisten Fällen keine sinnvollen Resultate, da die Motive zu stark voneinander abweichen.

Durch Vorgabe eines Schwellenwertes (zwischen 0,1 und 1,0) lässt sich außerdem bestimmen, wie stark die Ergebnisbilder vom Suchbild abweichen sollten. Je niedriger der Wert, desto größer die Differenzen. Hierbei ergeben sich bei einem Wert von 0,85 bis 0,95 die besten Treffer.

Die Attraktivität des Angebotes besteht in seiner Vielfalt – und in seiner Vollständigkeit und Aktualität. Denn auch die neuesten Digitalisate der Bayerischen Staatsbibliothek

stehen innerhalb weniger Tage für eine Bildähnlichkeitssuche zur Verfügung. Dafür sorgt eine automatisierte Infrastruktur im Hintergrund, die eng mit dem Backend des MDZ vernetzt ist. Neu digitalisierte Bücher werden umgehend erfasst, nach Bildinformationen ausgewertet und diese automatisch indiziert.

4 Fazit und Perspektiven

Mit rund 3000 Besuchern und mehr als 50.000 Seitenaufrufen pro Woche gehört die Bildähnlichkeitssuche zu den beliebtesten und erfolgreichsten Angeboten des Münchener Digitalisierungszentrums der Bayerischen Staatsbibliothek für Wissenschaft und Forschung. Mit der Erfassung aller Digitalisate ist die erstrebte maximale Ausbaustufe nun Ende 2016 erreicht. Die Bildähnlichkeitssuche richtet sich, wie dargestellt wurde, auf den immensen Reichtum an Bildern, Illuminationen, Holzschnitten, Graphiken, Zeichnungen und Emblemen, die in Handschriften, Inkunabeln und historischen Buchbeständen verborgen sind. „Verborgene“ deshalb, weil sie im Regelfall nicht einzeln und individuell erschlossen und in Katalogen erfasst sind. Für weite Teile der Geistes- und Kulturwissenschaften sind diese Bildbestände von größter Bedeutung, oft sogar wichtiger als die Texte selbst, in die sie eingebettet sind. Angesichts der großen und mit jedem neuen digitalisierten Buch weiter anwachsenden Zahl dieser visuellen Materialien – de facto handelt es sich um Massendaten – wird auch in langfristiger Perspektive ihre intellektuelle Erfassung und Katalogisierung nicht möglich sein.

Wie gezeigt wurde, verwendet die Suche ausschließlich automatisiert erkennbare, visuelle Merkmale der Bilder, konkret die Farb- und Kanteninformationen eines Bildes. Die Suche verzichtet damit vollständig auf herkömmliche Katalogdaten wie etwa bibliographische Informationen oder Schlagworte. Vereinfacht gesagt: Das Bild ist sein eigenes Katalogisat. Die neue Bildähnlichkeitssuche kann damit als Paradigma eines nicht-textbasierten Zugangs zum kulturellen Erbe gelten: die Auffindung und Segmentierung von Bildern in großen digitalen Textkorpora und die gezielte Ansteuerung der Bilder über die Ähnlichkeitssuche erfolgt ohne klassische „Metadaten“ ausschließlich auf der Basis der Eigenschaften des gesuchten Objektes selbst.

Auf diese Weise lassen sich Bilder quantifizieren, analysieren und vergleichen, ohne dabei bereits inhaltliche Konnotationen, künstlerische Intentionalität oder ein hermeneutisches Vorverständnis in den Vordergrund zu rücken. So können kunstwissenschaftlich und historisch relevante Zusammenhänge sichtbar werden, ohne dass ein „verstehendes“ Subjekt vorausgesetzt wird. Besonders deutlich wird dies mit Blick auf das Feature der Variation des Ähnlich-

keitsgrades der Bildsuche. Dieses hebt tendenziell die vertraute Dichotomie von maschineller Kalkulation einerseits und menschlicher Kreativität andererseits auf. Die Suche nach einer 100%-Ähnlichkeit führt nur auf identische Bilder, also beispielsweise auf die Verwendung desselben Motivs durch einen Künstler in mehreren Handschriften. Dies ist nichts, was der Experte im Regelfall nicht ohnehin schon weiß. Senkt man aber den vorgegebenen Ähnlichkeitsgrad etwa auf 95 % ab, zeigen die Resultate beispielsweise überraschende Motivvariationen oder formale und stilistische Übereinstimmungen bei ganz unterschiedlichen Bildinhalten. Dies sind maschinell generierte Ergebnisse, die ihrerseits unmittelbar ganz neue Forschungsfragen aufwerfen.

Indem die Ähnlichkeitssuche die Bildprogramme der digitalisierten Handschriften und Drucke ins Zentrum rückt, unterstützt sie zudem die „Emanzipation“ dieser visuellen Elemente aus ihrem textuellen Umfeld. Die dem Text ursprünglich und in der Intentionalität seiner Schöpfer nur „beigegebenen“, ihn „illustrierenden“ Elemente konstituieren sich so als „Kunstwerke“ und Forschungsobjekte eigenen Rechts. Wenn man will, könnte man sagen: Die Bildähnlichkeitssuche vollzieht den sogenannten „iconic turn“ gerade für die Objekte, deren textuelle Dominanz dem Bildlichen ursprünglich nur eine ornamentale Funktion zuweist: Handschriften, Inkunabeln und historische Drucke. Diese Emanzipation des Visuellen aus dem Kontext seiner textuell bestimmten Bedeutung setzt die Bildobjekte frei für eine neue, der Autonomie des Bildlichen verpflichteten Betrachtung und Analyse. Die visuelle Suche eröffnet damit ganz neue Perspektiven für eine digitale Bildwissenschaft.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Literatur

1. Bloomberg DS (1991) Multiresolution morphological approach to document image analysis. Proc. Int Conf Document Analysis Recognition (ICDAR 1991), 963–971
2. Bukhari SS, Shafait F, and Breuel TM (2011) Improved document image segmentation algorithm using multiresolution morphology. Proc. SPIE7874, Document Recognition and Retrieval XVIII
3. Ritendra D, Dhiraj J, Jia L, James ZW (2008) ages. Image Retrieval: Ideas, influences, and trends of the new age. ACM Comput. Surv. 40, 2, Bd. 5., S 60
4. Meiers T (2008) Vorrichtung und Verfahren zum Bestimmen eines Kanten-Histogramms, Vorrichtung und Verfahren zum Ablegen eines Bildes in einer Bilddatenbank, Vorrichtung und Verfahren zum Auffinden von zwei ähnlichen Bildern und Computerprogramm. Deutschland, DE 50 2008 001 725.9 und DE 50 2008 003 035.2