

Inhaltsbasierte Suche in Bilddatenbanken

Ullrich Köthe

Fraunhofer-Institut für Graphische Datenverarbeitung Rostock

D-18059 Rostock, Joachim-Jungius-Str. 9

Email: koethe@egd.igd.fhg.de

Zusammenfassung: Der vorliegende Artikel beschäftigt sich mit einer neuen Art intelligenter Informationssysteme, den Bilddatenbanken mit *inhaltsbasierter Suchoption*. Die inhaltsbasierte Suche gilt als vielversprechender Lösungsansatz für das Finden relevanter Daten in großen Datenbeständen. Die grundlegenden Konzepte in Bezug auf Bilddatenbanken werden beschrieben und anhand eines Modellbeispiels, einer Datenbank mit Brillengestellen, überprüft. Es zeigt sich, daß der Nutzen von inhaltsbasierter Suche entscheidend davon abhängt, daß das Retrievalsystem ähnliche Suchkriterien anwendet wie der Mensch. Aufgrund einfacher Experimente werden geeignete Kriterien, die auf Richtungshistogrammen und Skelettlinien beruhen, identifiziert. Die experimentelle Evaluierung des darauf aufbauenden Demonstrationssystems zeigt eine gute Übereinstimmung der Suchergebnisse mit den Erwartungen des Nutzers.

1. Einleitung

Die heutige Technologie gibt uns die Möglichkeit, riesige Mengen digitaler Bilddaten zu speichern, zu manipulieren und zu übertragen. Diese Entwicklung beruht unter anderem auf den ständig sinkenden Kosten für Speicherhardware und auf den stark verbesserten Zugriffsmöglichkeiten über weltweite Computernetze (Internet, WWW). Das bisher ungelöste Hauptproblem besteht dabei darin, in diesem gewaltigen Informationspotential diejenigen Daten zu finden, die für eine bestimmte Fragestellung relevant sind. Die Spannweite der möglichen Anfragen ist dabei sehr groß. So suchen beispielsweise Umweltwissenschaftler nach bestimmten Veränderungen in Satellitenbildern. Ärzte benötigen Röntgenaufnahmen, die eine bestimmte Besonderheit aufweisen. Kunden wählen in einem elektronischen Firmenkatalog Produkte aus (Teleshopping) usw.

Traditionell werden zur Suche in Bilddatenbanken Annotationen verwendet, die den Bildinhalt in Form von Wort- und Zahlenattributen kompakt wiedergeben. Diese Methode hat jedoch mehrere Nachteile:

- Ein Bild enthält normalerweise wesentlich mehr Informationen, als durch eine begrenzte Zahl von Annotationen ausgedrückt werden kann („Ein Bild sagt mehr als tausend Worte.“). Deshalb ist die Spanne der möglichen Fragestellungen sehr begrenzt.
- Annotationen müssen heute meist von Hand eingegeben werden. Dies verursacht einerseits hohe Kosten, andererseits bleiben die Annotationen von subjektiven Einschätzungen des Bearbeiters abhängig und sind schlecht objektivierbar.

Der wichtigste Ansatz zur Entwicklung intelligenter Suchmethoden besteht darin, mit Hilfe von Methoden der digitalen Bildanalyse den Bildinhalt selbst zu untersuchen. Diese neuen Methoden, deren Entwicklung weltweit zur Zeit stark vorangetrieben wird, werden unter dem Begriff „Inhaltsbasierte Suche“ (engl. „Content Based Retrieval“) zusammengefaßt.

Man unterscheidet drei grundlegende Prinzipien der inhaltsbasierten Suche in Bilddatenbanken:

on-line Bildanalyse: Hierbei entspricht jeder Anfrage ein Bildanalysealgorithmus, der nach Anwendung auf alle Bilder genau die gesuchten Bilder herausfiltert. Obwohl der Nutzer damit eine sehr große Flexibilität bei der Suche erhält, ist diese Methode für die meisten praktischen Anwendungen heute noch zu aufwendig.

automatische Annotation: Bei diesem Ansatz werden bei der Einfügung jedes Bildes automatisch Annotationen berechnet. Man kann dadurch wesentlich mehr Annotationen verwenden, als ein Mensch von Hand eingeben könnte. Standardisierte Algorithmen sichern außerdem die Vergleichbarkeit der Attribute. Da jede Suchanfrage mit Hilfe der Attribute formuliert werden muß, müssen diese sehr sorgfältig ausgewählt werden.

bildliche Schlüssel (pictorial indexes): Diese Methode versucht, die Vorteile der beiden vorhergehenden Methoden zu verbinden. Aus den ursprünglichen Bildern werden dabei (meist verkleinerte) Schlüsselbilder berechnet, in denen bestimmte Merkmale (z.B. Kanten) hervorgehoben und damit leicht zu

detektieren sind. Die Schlüsselbilder enthalten mehr Informationen als einfache Attribute, insbesondere bleibt der räumliche Kontext erhalten. Da umfangreiche Vorverarbeitungsschritte bereits durchgeführt wurden, ist der Rechenaufwand trotzdem viel geringer als bei der on-line Bildanalyse.

Aufgrund der Komplexität der Bildinformation kann man allerdings nicht erwarten, daß die inhaltsbasierte Suche eine exakte Übereinstimmung von Suchkriterien und Bildmerkmalen liefert. Statt dessen mißt man die *Ähnlichkeit* zwischen Suchkriterien und Bildinformation. Neuronale Netze und Fuzzy-Logik lassen sich, neben traditionellen Verfahren, hierzu erfolgversprechend einsetzen. Die Bilder mit der größten Ähnlichkeit werden dem Nutzer präsentiert, und dieser hat die letzte Entscheidung, welches er auswählen möchte. Eine Reihe neuer Anfragetypen werden dadurch möglich:

- beispielbasierte Suche:
 - "Suche Bilder, die einem Beispielbild (oder Bildausschnitt) ähnlich sind."
 - "Suche Bilder, die ein Objekt enthalten, das meiner Skizze ähnelt."
- Suche nach visuellen Merkmalen:
 - "Suche Bilder, die bestimmte Texturen und Farbverteilungen besitzen."
 - "Suche Bilder, die runde oder fünfeckige Objekte enthalten."
- iterative Suche:
 - Hat eine Anfrage kein befriedigendes Ergebnis erbracht, kann aus den besten Antworten eine neue Anfrage zusammengestellt und die Suche wiederholt werden, so daß man sich iterativ an die gesuchte Lösung herantastet. Dies erleichtert die Arbeit vor allem dann, wenn zwischen konkurrierenden Merkmalen ein Kompromiß gefunden werden muß.

Die Formulierung derartiger Anfragen stellt hohe Anforderungen an das Nutzerinterface, um effizient entsprechende Beispiele zu generieren, aus anderen Bildern zu extrahieren und zu modifizieren. Zusätzlich sollte die Möglichkeit zur Übersetzung anwendungsspezifischer Begriffe in den Merkmalsraum der Bildanalyse vorhanden sein, um Anfragen auch verbal formulieren zu können

Die meisten derzeit existierenden Prototypen realisieren die inhaltsbasierte Suche über automatische Annotationen. Das PHOTOBOK-System [1] konzentriert sich dabei auf die Ableitung optimaler Attribute für spezielle Fragestellungen, nämlich die Erkennung von Gesichtern, die Texturklassifikation mit Hilfe von WOLD-Merkmalen und die Formbeschreibung ebener Flächen durch Deformationsanalyse. QBIC [2], das derzeit einzige kommerziell verfügbare System, hingegen versucht, möglichst allgemein anwendbare Attribute zu bestimmen. Es werden daher insbesondere Farbhistogramme, einfache Texturmerkmale sowie Flächenmomente benutzt. Eine erste Anwendung der bildlichen Schlüssel wurde innerhalb des ART MUSEUM-Systems [3,4] verwirklicht. Spezielle verkleinerte Kantenbilder, die die Komposition des Bildes wiedergeben, werden hier zum Retrieval von Gemälden eingesetzt.

Die vorliegende Arbeit berichtet über ein Modellbeispiel, mit dessen Hilfe grundlegende Konzepte der inhaltsbasierten Suche demonstriert und überprüft wurden. Dieses Modellsystem ermöglicht die Suche nach Brillengestellen in einer entsprechenden Bilddatenbank, indem die gewünschte Brillenform durch eine Skizze oder ein Beispielbild beschrieben wird. Die Hauptschwierigkeit besteht hierbei darin, daß die vom System verwendeten Attribute dem Ähnlichkeitsempfinden eines Menschen sehr genau entsprechen müssen, da die Suchergebnisse sonst für den Nutzer kaum nachvollziehbar sind. Die relativ enge Eingrenzung des Anwendungsgebietes erleichtert dabei die Definition der Attribute sowie die Überprüfung der Ergebnisse unter verschiedenen Bedingungen.

Die Arbeit gliedert sich wie folgt: Im zweiten Abschnitt werden aufgrund psychologischer Untersuchungen geeignete Merkmale identifiziert. Entsprechende Algorithmen werden in Abschnitt 3 entwickelt. Abschnitt 4 beschäftigt sich schließlich mit der Formulierung von Anfragen und berichtet über Experimente zur Evaluierung der Suchmethoden.

2. Identifikation geeigneter Merkmale für die Brillenbeschreibung

Bereits eine oberflächliche Analyse zeigt, daß Menschen ein ausgeprägtes Empfinden für die Ähnlichkeit von Formen, in unserem Fall Brillengestellen, besitzen. Ein automatisches Retrievalsystem muß deshalb seine Ähnlichkeitskriterien so eng wie möglich an das menschliche Empfinden anpassen. Um den Begriff „menschliches Empfinden“ in diesem Zusammenhang genauer zu definieren, wurden zunächst eine Reihe einfacher psychologischer Tests durchgeführt. Grundlage dieser Tests war die Sammlung von Brillen in Abbildung 1. Auch wenn die Tests nur einen groben Einblick geben, zeigt sich, daß sie für das nachvollziehbare Funktionieren des Retrievals entscheidende Hinweise liefern. Ähnliche Tests müssen vermutlich Grundlage jeder inhaltsbasierten Suchumgebung sein.

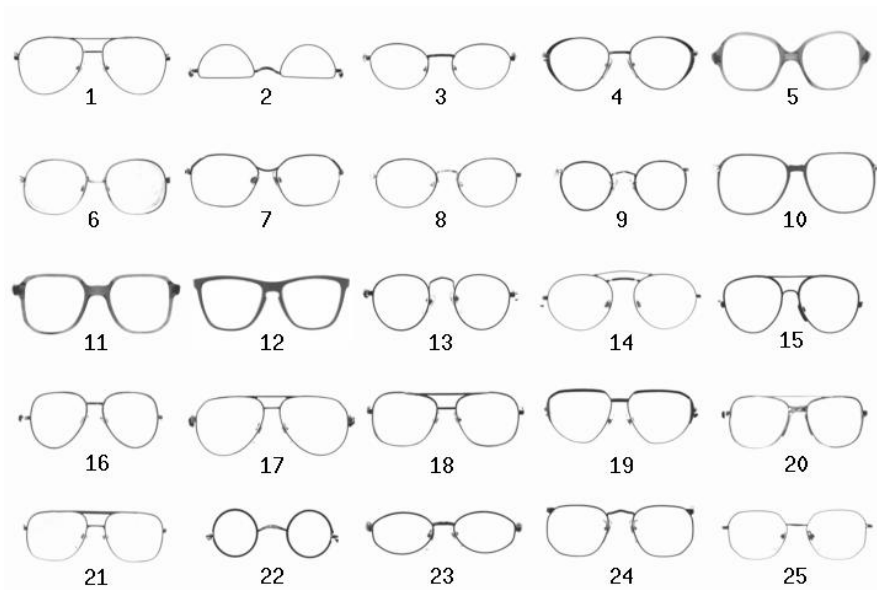


Abb. 1:
Brillen, die für
verschiedene Tests
verwendet wurden.
(Die Zahlen werden im
folgenden als
Referenzen verwendet.)

Zunächst ging es darum festzustellen, welchen Einfluß verschiedene Merkmale einer Brille auf die Beurteilung der Ähnlichkeit haben. Zusammenfassend ergab sich dabei folgendes:

- Wichtigstes Merkmal ist die Form des Brillenglases.
- Etwas geringer gewichtet sind die Dicke des Brillenrahmens, das Design des Nasenstegs sowie die Größe der Brille.
- Außer bei sehr auffälligem Design spielt der Schließblock (Ansatz der Bügel) eine geringere Rolle.

Farbe wurde als Kriterium zunächst nicht berücksichtigt, dies ist einer späteren Version des Programms vorbehalten.

Ausgehend von diesen Merkmalen ging es darum, grundlegende Klassifikationen für Brillen zu finden. Das entscheidende Merkmal für die Klassifikation ist offenbar die Zahl der Ecken der Grundform des Brillenglases. Daraus ergeben sich sechs Klassen: runde, ovale, drei-, vier- sowie fünf- und mehreckige Formen. Wichtig für die Ähnlichkeit ist auch, ob der Nasensteg aus ein oder mehr (meist zwei) Teilen besteht, woraus sich jeweils zwei Unterklassen ergeben. Die übrigen Merkmale zeigen keine deutliche Klassenbildung und werden deshalb nur kontinuierlich betrachtet.

Auf der Grundlage der Klassifikationen wurde nun die Ähnlichkeit von Brillen genauer erfragt. Ausgehend von typischen Klassenvertretern sollten die Versuchspersonen die anderen Brillen entsprechend ihrer Ähnlichkeit sortieren. Abb. 2 zeigt zwei Beispiele, in denen als Unterscheidungskriterium nur die Form des Brillenglases benutzt wurde. Die Position jeder Brille ergibt sich durch Mittelung ihrer Position in mehreren Test mit unterschiedlichen Versuchspersonen. Es ist zu erkennen, daß die Ähnlichkeit von der Richtung der „geraden“ Stücken des Randes sowie von der Schärfe der „Ecken“ bestimmt wird.

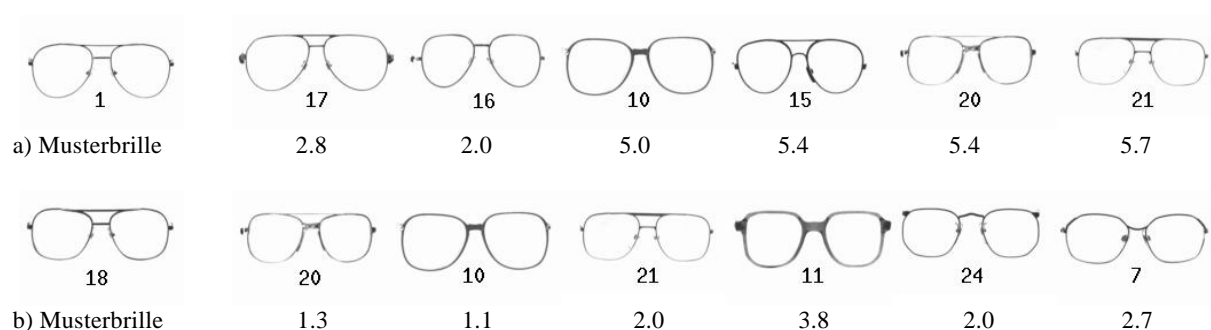


Abb. 2: Beispiele für die Einschätzung der Ähnlichkeit durch die Versuchspersonen. (links: Musterbrille; daneben: die sechs ähnlichsten Brillen aus Abb. 1 entsprechend ihrer mittleren Position bei allen Versuchspersonen. Die Zahlen unter den Bildern sind die Standardabweichungen dieser mittleren Position. Hohe Werte bedeuten, daß die Einschätzungen stark schwanken - im Extremfall um mehr als 5 Positionen in jede Richtung.

Aus den Tests ergibt sich folgende Liste von Merkmalen, die das Retrievalsystem berücksichtigen muß:

Brillenglas: Klassenzugehörigkeit (rund, oval, drei-, vier-, fünf-, mehreckig), Größe, Deskriptoren für die Details der Form (insbesondere Richtungen und Schärfe der Ecken)

Nasensteg: Klassenzugehörigkeit (ein Teil, mehrere Teile), Dicke, Breite, Position

Rahmen: mittlere Dicke insgesamt, mittlere Dicke in vier Richtungen (Nord, Süd, Ost, West)

Darüber hinaus wurde deutlich, daß verschiedene Versuchspersonen sehr unterschiedliche Gewichtungen der verschiedenen Merkmale vornehmen. Eine geeignete Nutzeranpassung muß durch das Retrievalsystem daher ebenfalls unterstützt werden.

3. Modellierung der Ähnlichkeit im Computer

Analyse des Brillenglases

Die Analyse des Brillenglases hat sich als das schwierigste Problem bei der Entwicklung erwiesen. Dies ist darauf zurückzuführen, daß sich die experimentell gefundenen Ähnlichkeitsbeziehungen nur schlecht mit herkömmlichen Deskriptoren wie Flächenmomenten [5] oder Fourierdeskriptoren [6] modellieren lassen. Offensichtlich ist die Information, die für Erkennung der subtilen Unterschiede zwischen konvexen Flächen (fast alle Brillengläser sind konvex) benötigt wird, über mehrere Deskriptoren verteilt und dadurch schwer zu extrahieren.

Innerhalb dieses Projekts wurde deshalb ein neuer Deskriptor entwickelt, der sich speziell für die Unterscheidung konvexer Flächen eignet. Er geht von der oben erwähnten Beobachtung aus, daß die Schärfe der Ecken und die Kantenrichtungen entscheidenden Einfluß auf die Unterscheidung haben. Wir nehmen an, daß das zu analysierende Gebiet in Parameterdarstellung vorliegt (der Parameter t ist der normierte Umfang):

$$\partial G = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}, \quad t \in [0,1]. \quad (1)$$

Der Winkel ψ zwischen der Tangente in einem Punkt des Randes und der x-Achse sowie die Länge dl eines differentiellen Kurvenstücks sind dann (mit $y'(t) = \frac{dy}{dt}$, $x'(t) = \frac{dx}{dt}$) gegeben durch:

$$\psi(t) = \begin{cases} \arctan \frac{y'(t)}{x'(t)}, & \text{falls } x'(t) \geq 0, y'(t) \geq 0 \\ \arctan \frac{y'(t)}{x'(t)} + \pi, & \text{falls } x'(t) < 0 \\ \arctan \frac{y'(t)}{x'(t)} + 2\pi, & \text{falls } x'(t) \geq 0, y'(t) < 0 \end{cases} \quad dl = \sqrt{x'(t)^2 + y'(t)^2}. \quad (2)$$

Wir definieren nun n Intervalle Ψ_i über dem Wertebereich von $\psi(t)$. Für ein konvexes Gebiet G ist die Funktion $\psi(t)$ monoton, so daß jedem Intervall Ψ_i eindeutig ein zusammenhängendes Intervall T_i über dem Definitionsbereich von $\psi(t)$ zugeordnet werden kann:

$$\Psi_i = \{\psi(t) : \psi_i \leq \psi(t) < \psi_i + \Delta\psi\}, \quad \psi_i = i \Delta\psi, \quad i = 0, \dots, n-1, \quad \Delta\psi = \frac{2\pi}{n} \quad (3)$$

$$T_i = \{t : \psi(t) \in \Psi_i\}$$

Mit Hilfe dieser Intervalle kann ein *Richtungshistogramm* für die Kurve definiert werden:

$$H[\Psi_i] = \int_{T_i} dl = \int_{T_i} \sqrt{x'(t)^2 + y'(t)^2} dt. \quad (4)$$

Das Richtungshistogramm mißt die Länge der Randkurve in jedem Winkelintervall $[\psi_i, \psi_i + \Delta\psi)$. Anschaulich bedeutet dies, daß H in einem Gebiet großer Krümmung (also an einer Ecke) einen kleinen Wert und in einem Gebiet kleiner Krümmung (Geradenstück) einen großen Wert besitzt. Deshalb ist es für die Beschreibung konvexer Flächen sehr geeignet.

Einige wichtige Eigenschaften des Richtungshistogramms sind:

- Es ist translationsinvariant.
- Es ist nicht rotationsinvariant. Diese Eigenschaft ist für die Brillenbeschreibung günstig, da Brillen eine definierte Lage besitzen.
- Es ist in Ψ periodisch (Periode 2π). Diese Eigenschaft kann genutzt werden, um rotationsinvariante Deskriptoren abzuleiten (beispielsweise die Beträge der Fourier-transformierten des Histogramms).
- Durch Verwendung des normierten Umfangs (d.h. $\sum_i H[\Psi_i] = 1$) ist das Histogramm skalierungsinvariant. Die absolute Größe der Brille, die natürlich ebenfalls interessant ist, wird als separates Merkmal durch die Quadratwurzel der Fläche des Brillenglases beschrieben.

In der Praxis ist die Berechnung des Richtungshistogramms sehr einfach, weil die Kurve normalerweise als Polygonzug, z.B. durch $\partial G = \{(x_k, y_k)^T, k = 0, \dots, m-1\}$, gegeben ist. In diesem Fall müssen anstelle der Differentiale Differenzen benutzt werden und es gilt:

$$H[\Psi_i] = \sum_{\{k: \Psi_k \in [\Psi_i, \Psi_i + \Delta\Psi]\}} l_k, \quad (5)$$

wobei Ψ_k die Richtung des k -ten Vektors des Polygonzuges und $l_k = \sqrt{(x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2}$ seine Länge sind.

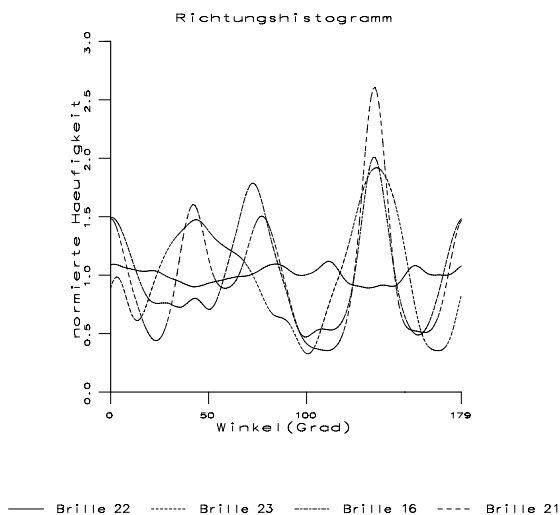


Abb. 3: Richtungshistogramme einiger Brillen
 Brille 22: fast rund => keine ausgeprägten Extrema
 Brille 23: oval => zwei Maxima
 Brille 16: dreieckig => drei Maxima
 Brille 21: viereckig => vier Maxima

Abb. 3 zeigt die (geglätteten) Richtungshistogramme für einige Brillen. Die Zuordnung der Brillen zu einer der Klassen rund, oval usw. ist jetzt relativ einfach. Im wesentlichen ergibt sie sich aus der Anzahl der Maxima im Richtungshistogramm. Häufig treten jedoch neben den deutlichen Maxima auch weniger deutliche auf. Typisches Beispiel ist eine fast runde, leicht ellipsenförmige Brille, die zwei kleine Maxima im Histogramm aufweist. In solchen Fällen kann die Brille in gewissem Grade mehreren Klassen zugeordnet werden. Ein einfacher *Fuzzy-Klassifikator* hat sich als das geeignete Werkzeug erwiesen, aus den Peaks des Histogramms solche graduellen Zugehörigkeitswerte zu berechnen.

Analyse des Brillengestells

Grundlage für die Beschreibung des Brillengestells ist das Skelett, welches mit Hilfe der Distanztransformation [7,8] bestimmt wird. Die Distanztransformation ermöglicht es außerdem, für jedes Pixel des Skeletts seinen Abstand vom Rand zu berechnen. Mit Hilfe eines Kantenverfolgungsalgorithmus lassen sich nun alle relevanten Pixel des Skeletts finden und über den gesamten Rahmen sowie innerhalb der vier Sektoren Nord, Süd, Ost und West (gemessen relativ zum Schwerpunkt eines Brillenglases) mitteln.

Lage, Größe und Anzahl der Nasenstege werden durch Analyse der Verzweigungspunkte des Skeletts ermittelt. Die Dicke ergibt sich wiederum als mittlerer Abstand der Skelettlinie des Nasenstegs vom Rand.

Definition der Ähnlichkeit

Bei der Definition der Ähnlichkeit müssen prinzipiell zwei Fälle unterschieden werden, nämlich der Vergleich kontinuierlicher Merkmale bzw. Merkmalsvektoren sowie der Vergleich von Klassenzugehörigkeiten.

Zur Berechnung der Ähnlichkeit kontinuierlicher Merkmale muß ein geeigneter Abstand zwischen den Merkmalen definiert werden. In vielen Fällen genügt der Euklidische Abstand, der für die Richtungshistogramme gegeben ist durch:

$$D^2(H_1, H_2) = \sum_i (H_1[\Psi_i] - H_2[\Psi_i])^2. \quad (6)$$

Eine Alternative beim Vergleich von Histogrammen bietet der symmetrische Abstand der zugehörigen Summenhistogramme (ähnlich dem Kolmogorov-Test):

$$D(H_1, H_2) = \sum_i \left| \sum_{k=0}^i H_1[\Psi_k] - \sum_{k=0}^i H_2[\Psi_k] \right|. \quad (7)$$

Die Ergebnisse sind bei beiden Varianten ähnlich, ausführlichere Untersuchungen zum Vergleich sind zur Zeit im Gange.

Mehrere voneinander unabhängige Distanzen lassen sich durch ein gewichtetes Mittel der Quadrate der Einzeldistanzen kombinieren:

$$D_{ges}^2(M_1, \dots, M_a) = \sum_{l=1}^a w_l D_l^2. \quad (8)$$

Die Gewichte w_l hängen dabei von der Varianz der Distanz D_l^2 sowie von ihrer Bedeutung für die Gesamtbeurteilung der Ähnlichkeit ab. Außerdem läßt sich der Ähnlichkeitsbegriff durch Modifikation der Gewichte leicht an verschiedene Nutzeranforderungen anpassen. Dies wird weiter unten ausführlicher diskutiert.

Um die Ähnlichkeit von Klassenzugehörigkeiten zu testen, hat man zwei prinzipielle Möglichkeiten. Im einfachsten Fall berücksichtigt man nur Objekte, die innerhalb der gleichen Klasse liegen wie das Suchprädikat. Dies führt jedoch häufig zu einer zu starken Einschränkung des Suchraumes. Hier ist es besser, auch zwischen den Klassen eine „Cluster-Distanz“ zu definieren (z.B. mit Hilfe der Euklidischen Distanz der Clusterzentren). Diese kann dann mit einem geeigneten Gewicht ebenfalls in die Gesamtdistanz D_{ges} eingehen.

Die Unterscheidung von klassenbasierten und kontinuierlichen Merkmalen gestattet darüber hinaus eine einfache Art der Indizierung größerer Datenbestände. Die Suche in der Datenbank wird dann in zwei Schritten durchgeführt. Zuerst werden mit wenigen klassenbasierten Merkmalen diejenigen Brillen identifiziert, die im gleichen oder in benachbarten Clustern liegen wie das Suchprädikat. Nur bei diesen ist danach der viel aufwendigere Vergleich aller Merkmale notwendig, was zu einer deutlichen Beschleunigung der Suche führt.

4. Anfrageformulierung und Experimente

Zwei grundlegende Anfrageparadigmen wurden bisher implementiert und getestet: Suche mit Hilfe eines Beispiels sowie Suche aufgrund einer Skizze. Die Eingabe der Skizze erfolgt dabei mit einem speziellen Editor, der die intuitive Festlegung der gewünschten Brillenform auf der Basis von Splines gestattet. Die Werte für die übrigen Merkmale werden durch Schieberegler und Buttons festgelegt. Bei der Beispielsuche wird im einfachsten Falle eine Brille der Datenbank als Vergleichsobjekt festgelegt. Wählt man mehrere Beispiele aus, kann jedes Beispiel Zielmerkmale aus einer anderen Merkmalsgruppe liefern, so daß der Nutzer aus mehreren teilweise befriedigenden Antworten eine verfeinerte Fragestellung ableiten kann.

Außerdem gibt es die Möglichkeit, durch Modifikation der Gewichte die Suche an Nutzerpräferenzen anzupassen. So könnte beispielsweise ein Nutzer besonderen Wert auf kleine Brillen legen, während ein anderer auf keinen Fall einen mehrteiligen Nasensteg akzeptiert. Dieses Feature läßt sich während der Testphase auch dazu verwenden, die Suche nach jedem Attribut getrennt zu testen, indem die Gewichte der anderen Merkmale null gesetzt werden.

Die Testmethodik ist an Techniken aus [2] angelehnt. Die Ergebnisse der automatischen Suche werden mit dem entsprechenden Experiment verglichen. Die Güte der Suche ergibt sich aus der mittleren bzw. der maximalen Differenz zwischen der experimentell bestimmten Sollposition der 6 ähnlichsten Brillen sowie deren

tatsächlicher Position beim automatischen Retrieval. Folgende Tests wurden durchgeführt (jeweils für Beispielsuche und Suche mit Skizze):

- Test der Ähnlichkeit aufgrund der Klasseninformation allein
- Test der Ähnlichkeit des Richtungshistogramms innerhalb der selben und benachbarter Klassen
- Test der Rahmenmerkmale
- Test der Nasenstegmerkmale
- Tests mit verschiedenen Merkmalskombinationen

Die Tests aufgrund der Klasseninformationen zeigen, daß das System die Klassenzugehörigkeiten korrekt erkannt hat. Vergleicht man zusätzlich die Richtungshistogramme, so daß auch die Ähnlichkeit innerhalb der Klassen gemessen wird, ergeben sich die Resultate in Abb. 4. Die Musterbrillen sind die selben wie in Abb. 2, die Reihenfolge in Abb. 2 ist also die Soll-Reihenfolge. Sollposition und tatsächliche Position weichen im Mittel um 1.6, höchstens um 3 Plätze voneinander ab. Diese Abweichungen entsprechen etwa den Standardabweichungen bei den Experimenten, d.h. die Abweichungen des Retrievalsystems sind nicht größer als die Abweichungen der Versuchspersonen untereinander. Damit ist eine wichtige Bedingung für die praktische Anwendbarkeit des Verfahrens erfüllt.

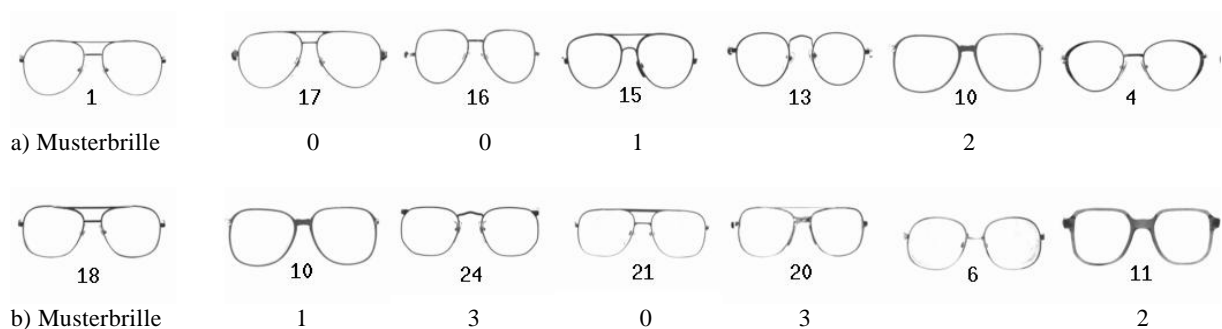


Abb. 4: Ergebnis des automatischen Retrievals der Brillenform für die selben Musterbrillen wie in Abb. 2 (links: Musterbrillen; daneben: die sechs ähnlichsten Brillen. Die Zahlen geben die Abweichungen zur Sollposition an, wenn die selbe Brille in Abb. 2 enthalten war. Die noch fehlenden Vergleichsbrillen befinden sich auf den Positionen 7 bzw. 8)

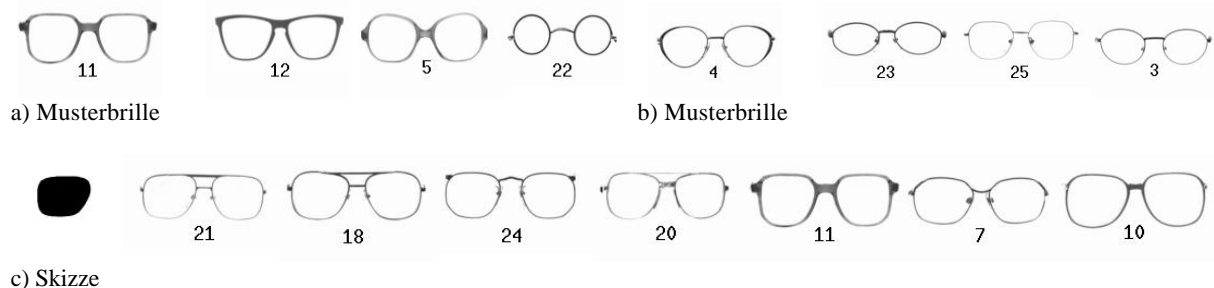


Abb. 5: Beispiele für Suche aufgrund der Rahmendicke (a), der Form des Nasenstegs (b) und mit Hilfe einer Skizze (c)

Die Abbildung 5 zeigt Resultate für weitere Merkmale sowie für die Suche mit Hilfe einer Skizze. Auch hier ist zu erkennen, daß die Antworten des automatischen Suchsystems den Ergebnissen der Experimente sehr nahekommen.

Das System wurde in C++ auf der Basis der objektorientierten Datenbank VERSANT auf einer SparcStation 10 implementiert. Die Antwortzeit des Systems beträgt etwa 0.4 ms pro Brille für einen vollständigen Merkmalsvergleich. Der Vergleich der Klassenzugehörigkeiten allein benötigt 0.1 ms. (Dabei sind die Retrievalzeiten der Merkmale in der Datenbank nicht enthalten, weil die entsprechenden Objekte im Hauptspeicher gecached werden.) Einfache lineare Suche ist demnach bis etwa 1000 Brillen schnell genug für die interaktive Arbeit. Bei klassenbasierter Indizierung verkürzt sich die Retrievalzeit etwa auf die Hälfte, wenn durch die Vorauswahl nur noch bei 1/3 der Brillen eine vollständige Suche notwendig ist.

5. Ergebnisse und Ausblick

In der vorliegenden Arbeit wurde ein allgemeines Konzept zur inhaltsbasierten Suche in Bilddatenbanken entwickelt und an einem Beispiel getestet. Die Tests zeigen, daß die inhaltsbasierte Suche ein leistungsfähiges, für den Nutzer intuitiv verständliches Mittel zum Auffinden visueller Informationen ist. Von entscheidender Bedeutung ist dabei allerdings, daß die vom System verwendeten Merkmale und Ähnlichkeitsmaße eng an die entsprechenden, experimentell bestimmten Begriffe von Menschen angelehnt sein müssen. Außerdem müssen Anpassungsmöglichkeiten an die Variabilität und Subjektivität dieser menschlichen Begriffe geschaffen werden.

Die Suche mit Hilfe einer Skizze eignet sich nach unseren Erfahrungen besonders, um am Anfang der Suche erste Antworten zu generieren. Danach ist es am einfachsten, sich aus den gefundenen Beispielen zusammenzustellen, so daß die Anfrage immer exakter wird. Auf diese Weise kann man sich von einer groben Skizze sehr schnell iterativ an eine Lösung herantasten.

Zur Zeit erforschen wir, wie diese Konzepte auf kompliziertere Bilder, z.B. normale Fotografien, ausgedehnt werden können. Die wichtigsten Probleme sehen wir bei der Segmentierung der Bilder sowie bei der Bestimmung geeigneter, anwendungsübergreifender Merkmale. Auch Fragen des User-Interfaces werden eine sehr große Rolle spielen, weil die bisherigen Werkzeuge für die Anfragegenerierung und für die Visualisierung von Ähnlichkeiten nicht mehr ausreichen. Innovative Methoden zur Indizierung vieldimensionaler Merkmalsräume werden mit dem Anwachsen der Datenbestände immer wichtiger. Die Vielzahl der potentiellen Anwendungen, von elektronischen Katalogen für das Teleshopping über schnellen Zugriff auf medizinische Bilddaten bis zur problembasierten Suche in Satellitenbildarchiven, versprechen jedoch einen großen „return-on-investment“ für die notwendige Forschungsarbeit.

6. Literaturverzeichnis

- [1] A. Pentland, R. Picard, S. Sclaroff: „Photobook: Tools for Content-Based Manipulation of Image Databases“, in: SPIE Conf. on Storage and Retrieval in Image and Video Databases II, No. 2185, Feb. 1994
- [2] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, W. Equitz: „Efficient and Effective Querying by Image Content“, in: J. of Intelligent Information Systems, 3 (1994)
- [3] T. Kato: „Database architecture for content-based image retrieval“, in: SPIE Conf. on Image Storage and Retrieval Systems, No. 1662, Feb. 1992
- [4] K. Hirata, T. Kato: „Query by Visual Example - Content-based Image Retrieval“, in: A. Pirotte, C. Delobel, G. Gottlob (Eds.): Advances in Database Technology - EDBT'92, Springer Lecture Notes in Computer Science, Vol. 58, 1992
- [5] C.-H. Teh, R.T. Chin: „On Image Analysis by the Method of Moments“, PAMI 10(4), pp. 496-513, 1988
- [6] N. Kiryati, D. Maydan: „Calculating Geometric Properties from Fourier Representation“, Pattern Recognition 22(5), pp. 469-475, 1989
- [7] A. Montanvert: „Medial Line: Graph Representation and Shape Description“, in: Proc. 8th ICPR 1986
- [8] F.L. Krause, H. Jansen, N. Luth: „Neue Methoden der automatischen Zeichnungsinterpretation“, in: ZWF CIM, 88. Jahrgang, S. 589-592, 1993