Wozu „Inhaltsorientierte Suche“?

Vorlesung Bildverarbeitung 2, Teil 1
Wintersemester 2001/2002
Ullrich Köthe, FB Informatik, Uni Hamburg

Literatur (1)

In der Bibliothek:

- A. del Bimbo: Visual Information Retrieval, 1999
  Einführung in die grundlegenden Verfahren
- S. Santini: Exploratory Image Databases, 2001
  umfassende theoretische Behandlung fortgeschrittener Verfahren
- Y. Rubner, C. Tomasi: Perceptual Metrics for Image Database Navigation, 2001
  Theorie und Experimente zu Distanzen (bes. *earth mover distance*)

Im Handel:

- C. Faloutsos: Searching Multimedia Databases by Content, 1996
  ähnlich del Bimbo
- A. Smeulders, R. Jain (eds.): Image Databases and Multi-Media Search, 1997 (Sammelband)
- V. Castelli, L. Bergman: Image Databases, 2001 (noch nicht erschienen)

Literatur (2)

zahlreiche Artikel in Zeitschriften und Konferenzen:

Smeulders et al.: „Content Based Image Retrieval at the End of the Early Years“, Trans. Pattern Analysis and Machine Intelligence, 22(12), 1349-1380, 2000

Rui et al.: „Image Retrieval: Current Techniques, Promising Directions, and Open Issues“, J Visual Communication and Image Representation 10, 39-62, 1999

Konferenz: Storage and Retrieval in Image and Video Databases

im Internet:

Linkliste zu Firmen und Forschungsgruppen: http://www-student.informatik.uni-bonn.de/~gerdes/CBIR/

Überblick über 39 Systeme (Oktober 2000): http://ftp.cs.uu.nl/pub/RUU/CS/techreps/CS-2000/2000-34.pdf

Umfassender *state-of-the-art report* (Januar 1999) http://www.unn.ac.uk/iidr/report.html

Anwendungen von Bilddatenbanken

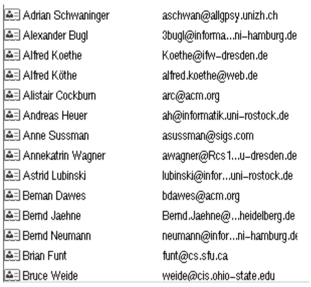
- **Medizin, Forschung (bildgebende Verfahren)**
- **Biometrie (Gesicht, Iris und Retina, Fingerabdruck)**
- **Bildarchive (Presse und Buch, Multimedia, Werbung, Kunst, Geschichte, Denkmalpflege, Fotoalbum)**
- **elektronische Kataloge (Handel, Makler, Reisebüros)**
- **Fernerkundung, Kartographie, GIS**
- **Industrie (Teilekataloge, optische Prüfung)**
- **WWW**
- **Video on Demand**

⇨ **Digitalisierung der Bestände in den letzten 10 Jahren**

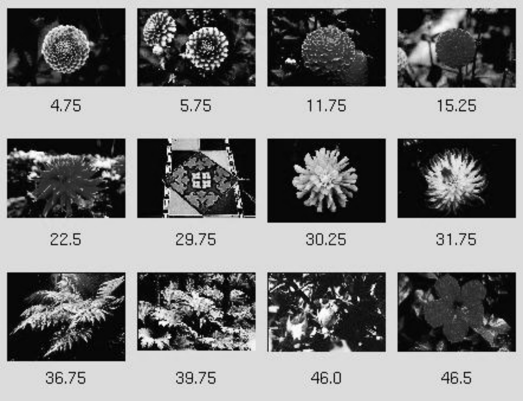
Traditionelle Suchtechniken: Browsing

Parallelverarbeitung des Auges

automatisches
Ignorieren der
false positives



Adrian Schwaninger	aschwan@allgpsy.unizh.ch
Alexander Bugl	3bugl@informa...ni-hamburg.de
Alfred Koethe	Koethe@ifw-dresden.de
Alfred Köthe	alfred.koethe@web.de
Alistair Cockburn	arc@acm.org
Andreas Heuer	ah@informatik.uni-rostock.de
Anne Sussman	asussman@sigs.com
Annekatrin Wagner	awagner@Rcs1...u-dresden.de
Astrid Lubinski	lubinski@infor...uni-rostock.de
Beman Dawes	bdawes@acm.org
Bernd Jaehne	Bernd.Jaehne@...heidelberg.de
Bernd Neumann	neumann@infor...ni-hamburg.de
Brian Funt	funt@cs.sfu.ca
Bruce Weide	weide@cis.ohio-state.edu

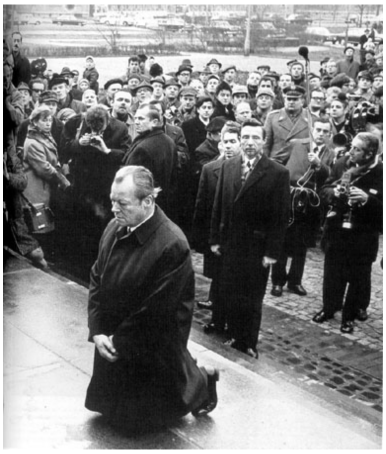


Traditionelle Suchtechniken: assoziierte textuelle Beschreibung (1)

- **Textannotationen sind notwendig**

  „Willi Brandts Kniefall vor dem Denkmal für die Opfer des Warschauer Aufstandes , 1970“

- **Text hat unmittelbar semantische Bedeutung**
- **Information jenseits des reinen Bildinhaltes**
- **Algorithmen vorhanden**



Traditionelle Suchtechniken: assoziierte textuelle Beschreibung (2)

- **Freier Text**
  - Bildunterschriften, umgebender Text
  - ungebundenes Vokabular, Index einfach zu erstellen
  - Methoden der Volltextsuche
- **Schlüsselwörter**
  - gezielte Zuordnung bei der Erfassung
  - gebundenes Vokabular vorteilhaft
  - teilweise automatisch: Datum, Autor etc.
- **Systematische Klassifikation**
  - vorgegebene Schlüsselwörter mit Definition
  - Hierarchie, Querbezüge, Synonyme
  - Erstellung ist schwieriges Forschungsprojekt!

Traditionelle Suchtechniken: assoziierte textuelle Beschreibung (3)

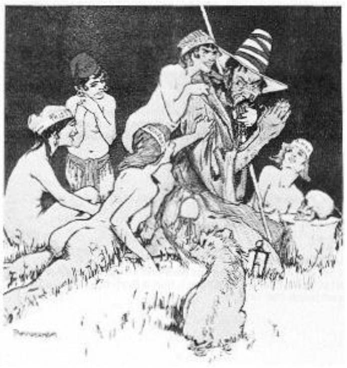
- **jede kommerzielle Bilddatenbank hat Annotationen**
- **teilweise standardisiert:**
  - AAT - Art & Architecture Thesaurus (Renselaer Polytechnic Institute, Getty Information Institute, 1980-heute)
  - ICONCLASS Iconographic Classification for Art History (Uni Leiden und Utrecht, 1960-heute)
  - TELCLASS - TV Klassifikation (BBC, 1987)
  - Opitz-Codierung für Maschinenteile (Opitz et al., 1969)
  - Wiener Klassifikation der Bildmarken (WIPO - World Intellectual Property Organization, 1992)
  - traditionelle heraldische Beschreibungen
  - Pressearchive ?

Traditionelle Suchtechniken: assoziierte textuelle Beschreibung (4)

- **ICONCLASS:**
  - 40 Jahre Entwicklung, 24 000 Definitionen, 5 000 Querverweise, 17 Bände (1973-1985), jetzt computerisiert
  - Anfragebeispiel: Wie wurde die Verkündigungsszene typischerweise im 17. Jh. in Holland gemalt - mit/ohne Engel, in der Kirche/anderem Raum?
  - Zahlen-/Buchstabencode, in eckigen Klammern: Bilddetail
- **AAT - Art & Architecture Thesaurus**
  - 120 000 Begriffe (Hauptbegriffe mit Definition und Synonymen)
  - Querverweise (assoziativ)
  - Hierarchie: 7 Kategorien (*facets*) mit 33 Subkategorien
    - *Objects, Materials, Activities, Agents (people, organizations), Styles and Periods, Physical Attributes (properties, conditions, color,...) Associated Concepts (cultural background)*
- **Wiener Klassifikation: ca. 1500 Kategorien, 3-stufige Hierarchie**

Beispiel: ICONCLASS

Schlagworthierarchie



1 = *Religion und Magie*
(Schlüsselwörter: übernatürlich)

11 = *Christentum*
(SW: + Religion,
Querverweis: 7 - Bibel)

11H = *männlicher Heiliger*
(SW: + Heiliger)

11H(Antony Abbot) = *Hl. Antonius*
(SW: + Buch, Glocke, Schwein, Flamme,
T-förmiger Stab, Abt, Eremit)

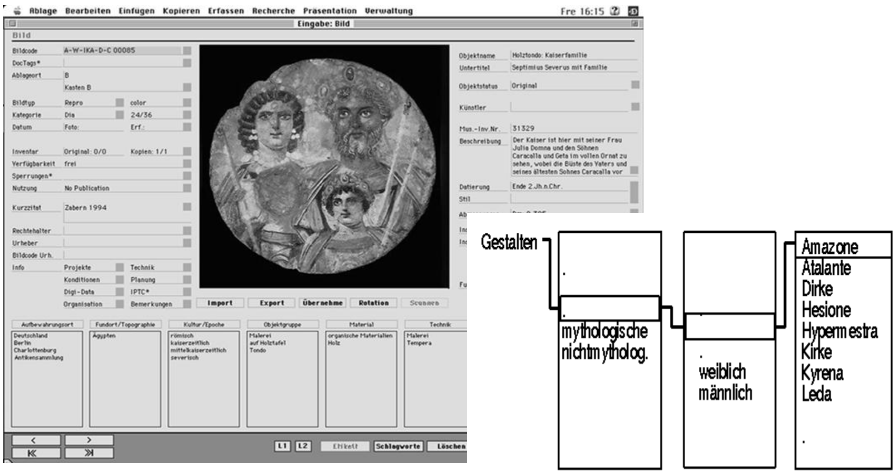
11H(Antony Abbot)3 = *Andacht des Hl.*

11H(Antony Abbot)35 = *die Versuchungen des Hl. Antonius*

11H(Antony Abbot)352 = *die erotischen Visionen des Hl. Antonius*
(SW: + Visionen, nackt, Frau, Sinnlichkeit, Kruzifix, zeigen, beten, vertreiben)

Beispiel: Image Finder Software

Auswahl aus Schlagworthierarchie


Ablage
Bearbeiten
Einfügen
Kopieren
Erfassen
Recherche
Präsentation
Verwaltung
Fre 16:15
Eingabe: Bild
Bild
Import
Export
Übernehme
Rotation
Scannen
L1
L2
Etikett
Schlagworte
Löschen
Gestalten
mythologische
nichtmythologische
weiblich
männlich
Amazone
Atalante
Dirke
Hesione
Hypermestra
Kirke
Kyrena
Leda

Bildcode A-W-IKA-D-C 00085
DocTags*
Ablageort B
Kasten B
Bildtyp Repro color
Kategorie Dia 24/36
Datum Foto: Erf.:
Inventar Original: 0/0 Kopien: 1/1
Verfügbarkeit frei
Sperrungen*
Nutzung No Publication
Kurzzitat Zabern 1994
Rechtehalter
Urheber
Bildcode Urh.
Info Projekte Technik
Konditionen Planung
Digi-Data IPTC*
Organisation Bemerkungen

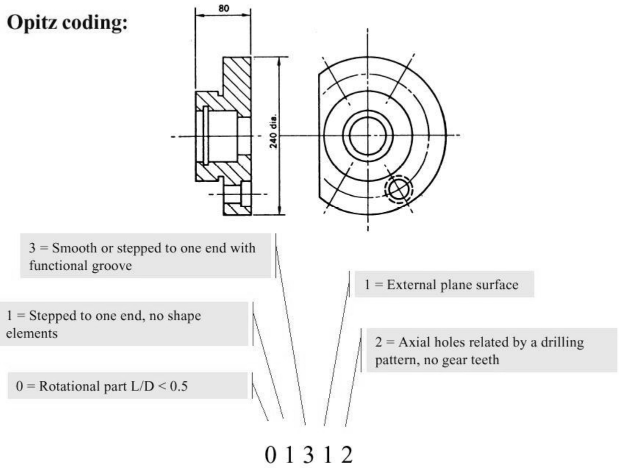
Objektname Holztondo: Kaiserfamilie
Untertitel Septimius Severus mit Familie
Objektstatus Original
Künstler
Mus.-Inv.Nr. 31329
Beschreibung Der Kaiser ist hier mit seiner Frau Julia Domna und den Söhnen Caracalla und Geta im vollen Ornat zu sehen, wobei die Büste des Vaters und seines ältesten Sohnes Caracalla vor
Datierung Ende 2.Jh.n.Chr.
Stil

Aufbewahrungsort	Fundort/Topographie	Kultur/Epoche	Objektgruppe	Material	Technik
Deutschland Berlin Charlottenburg Antikensammlung	Ägypten	römisch kaiserzeitlich mittelkaiserzeitlich severisch	Malerei auf Holztafel Tondo	organische Materialien Holz	Malerei Tempera

---

Traditionelle Suchtechniken: assoziierte textuelle Beschreibung (5)

- **Opitz-Code: 5-stelliger Zahlencode für Maschinenteile**


Opitz coding:
80
240 dia.
3 = Smooth or stepped to one end with functional groove
1 = External plane surface
1 = Stepped to one end, no shape elements
2 = Axial holes related by a drilling pattern, no gear teeth
0 = Rotational part L/D < 0.5
0 1 3 1 2

Traditionelle Suchtechniken: assoziierte textuelle Beschreibung (6)

- **Nachteile von Schlagwortsuche**
  - hoher Aufwand: Erstellung der Klassifikation (ICONCLASS: 40 J.), Zuordnung der Klassen (AAT: 7-40 min), Schulung der Nutzer
  - nicht alles kann beschrieben werden: z.B. Gesichter, Formen, Stimmungen, Details
    Beispiel: Maschinenbaufirma konnte Duplikate nicht identifizieren
  - schlechte Reproduzierbarkeit der Klassenzuordnung (Dyson 1992)

⇨ **Ergänzung durch visuelle Anfragen**

⇨ **Beziehungen zwischen Bild und Text herstellen**

⇨ **Schlagworte (teil)auto-matisch zuordnen**

⇨ **Lernen von Beschreibungen während der Benutzung**

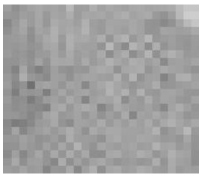
Untersuchung des Nutzerverhaltens

- **Studien über Anfragetypen an Bilddatenbanken:**
  - Pressearchive: Anfrage nach
    - speziellem Bild, Objekt oder Person (überwiegend)
    - Metadaten (Urheber, Rechteinhaber etc.)
    - Hintergrundinformation (spezielle Ereignisse etc.)
    - allgemeinen Themen, Abstraktionen, Stimmungen (seltener)
  - Nutzergruppen mit unterschiedlichen Bedürfnissen
  - Studien in anderen Anwendungsgebieten mit ähnlichen Ergebnissen: überwiegend schlagwortbasierte Anfragen

⇨ **Reichen Schlagworte aus, oder passen sich die Nutzer nur den Einschränkungen an ?**

Was bedeutet ein Bild? (1)

- **„Ein Bild sagt mehr als tausend Worte“**
  - ⇨ ja, aber es verbraucht soviel Speicher wie 10 000 Worte
- **Bilder haben viele Bedeutungsebenen**
  - – Meta-Ebene: „Suche Bilder, die 1970 entstanden sind!“, „Suche Bilder der Größe 1024x786!“
  - – Pixelebene: „Suche Bilder, bei denen der Punkt (12, 20) den Grauwert 255 hat!“ (z.B. Suche nach Fehler in der Kamera)
  - – Gruppen von Pixeln: „Suche Bilder, die ein vergleichbares Pixelmuster enthalten!“ (Suche nach JPEG-Artefakten)



Was bedeutet ein Bild? (2)

- **Bilder haben viele Bedeutungsebenen (Forts.)**
  - – Gesamtheit der Farben: „Suche Bilder, die typische Farben eines Sonnenuntergangs haben!“ (Suche nach Farbstimmung)
  - – Beachte Trennung Hintergrund/Vordergrund, Zentrum/Rand

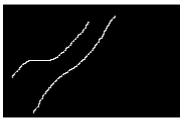
Was bedeutet ein Bild? (3)

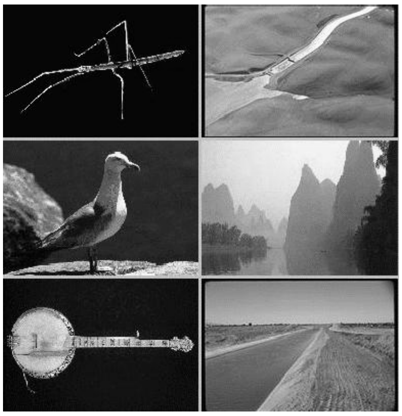
- **Bilder haben viele Bedeutungsebenen (Forts.)**
  - Objekteebene: „Suche Bilder, die ein Flugzeug enthalten!“
    „Suche Bilder, die eine Wiese enthalten!“
    (Textur)

N43
Lufthansa

Was bedeutet ein Bild? (4)

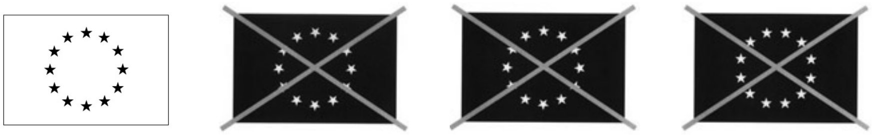
- **Bilder haben viele Bedeutungsebenen (Forts.)**
  - Formmerkmale: „Suche Bilder mit runden Objekten!“
  - Kantenmerkmale: „Suche Bilder mit Kanten gemäß der Skizze“ (Beispiel: Curvelet-based Image Retrieval, Lei et al. 97)





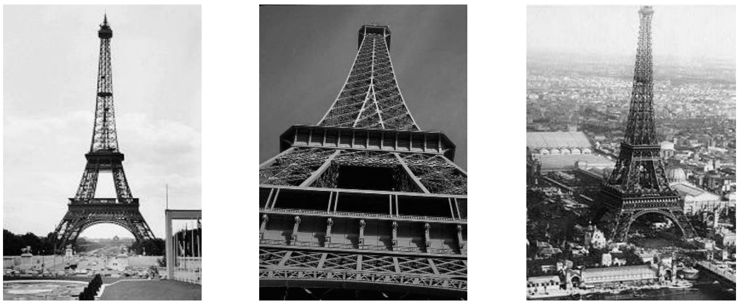
Was bedeutet ein Bild? (5)

- **Bilder haben viele Bedeutungsebenen (Forts.)**
  - räumliche Beziehungen: „Suche Bilder mit gelben Sternen auf einem Kreis vor blauem Hintergrund“
  - Heraldische Beschreibung: ein Kranz von 12 goldenen fünfzackigen Sternen, deren Spitzen sich nicht berühren, auf azurblauem Hintergrund
  - geometrische Beschreibung: Feld 1x1.5, Kreisradius 1/3, Sterne bei k*30°, Radius 1/18, alle Spitzen nach oben



Was bedeutet ein Bild? (6)

- **Bilder haben viele Bedeutungsebenen (Forts.)**
  - Ebene der 3D-Welt: räumliche Relationen in 3D, Perspektive „Suche Bilder, wo ein Schiff hinter dem Leuchtturm vorbeifährt!“ „Suche Bilder des Eiffelturms, unabhängig von der Perspektive!“



Was bedeutet ein Bild? (7)

- **Bilder haben viele Bedeutungsebenen (Forts.)**
  - Interpretationsebene: „Suche Bild mit dem Kniefall von Willi Brandt!“, „Suche Bild, das ein Kernkraftwerk (und kein Kohlekraftwerk) enthält!“
  - Ebene der Beziehungen zwischen Bildern: „Suche Bilder von New York vor und nach der Katastrophe!“
- **die Ebenen sind relativ unabhängig voneinander**

⇨ **kombinierte Anfragen**
  - „Suche Bild mit Sonnenuntergangsstimmung, das keinen Sonnenuntergang zeigt!“
  - „Suche *aerial view* des Eiffelturms!“ (also nicht „vom Turm herunter geschaut“, sondern „vom Flugzeug auf den Turm“)

Was bedeutet ein Bild? (8)

- **„Ein Bild sagt mehr als tausend Worte“**
  ⇨ ein Bild kann viele Bedeutungen haben
- **Bedeutung ergibt sich oft aus der Anfrage**
  ⇨ Unterschied zu traditionellen Datenbanken: feste Bedeutung
- ***Sensory Gap:***
  - Informationsverlust bei Projektion von 3D auf 2D
  - zufällige Aspekte der Abbildung (Beleuchtung, Perspektive)

  ⇨ wichtige Informationen sind verloren oder überlagert
- ***Semantic Gap:***
  - einzelne Pixel haben (meist) keine Bedeutung, sondern nur Gruppen von Pixeln und deren Relationen

  ⇨ Semantik nicht direkt aus den Daten ableitbar

Multimodalität

- **Bilder sind selbst multimodal:**
  - Farbkanäle, Multispektralbilder, bildgebende Verfahren
  - Bedeutungsebenen auch als Modi interpretierbar
- **Bilder oft mit anderen Daten kombiniert**
  - Text (Bildunterschriften, Schlagworte, umgebender Text)
  - WWW: Stellung des Bilds in der Seite, Vernetzung
  - Video:
    - Texteinblendungen (Nachrichten, Abspann)
    - Bewegung, Schnitte
    - Ton (Sprache, Geräusche, Musik - selbst multimodal)

⇨ **Kombination der Modi, *Sensorfusion***

Folgerungen für die Bildsuche (1)

- **allgemeine Lösung: Übergebe einen Algorithmus, der genau die gewünschten Bilder liefert, an die DB**
  - zu langsam (große Datenmengen!)
  - zu kompliziert, sogar für Experten (da keine 1-1 Entsprechung zwischen Algorithmus und Semantik)
  - ⇨ Ableiten von leichter zugänglichen Daten aus den Rohdaten, z.B. Histogramme, Signaturen, Repräsentationen
- **Einteilung der Anwendungen:**
  - *Closed World* (enge Domäne): relative eindeutige Bedeutung, z.B. Gesichter, Fingerabdrücke, Logos ⇨ traditionelle Erkennungsaufgabe
  - *Open World* (breite Domäne): keine eindeutige Bedeutung, z.B. Bildarchive, Videodatenbanken ⇨ neue Methoden

Folgerungen für die Bildsuche (2)

- **„inhaltsorientierte Suche“**
  - Suche, die viele semantische Ebenen unterstützt, nicht auf vorbestimmte Interpretation begrenzt
- **„*User in the loop*“**
  - Benutzer der Datenbank als integraler Bestandteil des Gesamtsystems
  - Semantik vom (anpassungsfähigen) Benutzer induziert
- **Ähnlichkeitssuche**
  - keine exakte Übereinstimmung von Anfrage und Ergebnis
- **Exploration**
  - iterative Suche, Relevanzfeedback, Modifikation der Anfrage („Übersetzung der Semantik“), spezielle *User Interfaces*

Beispiele für CBR-Systeme (1)

- **etwa 40 Systeme, einige kommerziell**
  - teilweise *closed world*, aber oft allgemein
  - fast alle bieten Farbsuche
  - oft Suchmöglichkeit nach Texturen und Form
  - Suche nach Komposition, Kanten etc. oft unbrauchbar
- **Geschichte:**
  - erste Versuche in den 70er Jahren (bes. Fingerabdrücke)
  - Erstes größeres System: Kato et al. ab Mitte der 80er (Kantenbilder und Farbeindruck)
  - QBIC (IBM): Anfang der 90er (grundlegende Verfahren der Farb- und Textursuche, Formsuche, Kommerzialisierung)
  - seitdem stetige Zunahme der Aktivitäten

Beispiele für CBR-Systeme (2): ART MUSEUM, TRADEMARK

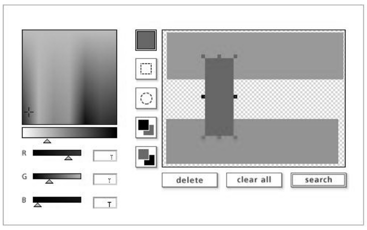
- **Kato et al. (Tsukuba Science City, Japan), 1985-92**
- **erstes großes Projekt zur visuellen Suche**
  - Bildmarken, Gemälde
- **Einführung grundlegender Begriffe:**
  - CBIR: *Content Based Image Retrieval*
  - QVE: *Query by Visual Example* (Skizze oder Beispielbild)
- **Bildbeschreibung durch**
  - *image abstraction* (vergröbertes Kantenbild, *graphical feature parameters*)
  - Farbeindruck (Lernen von nutzerbezogenen, subjektiven Beschreibungen - „modern“, „klassisch“, „warm“, „kalt“)

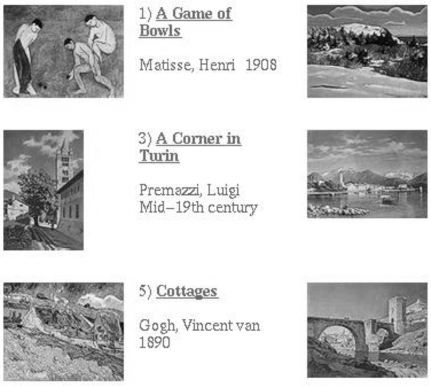
Beispiele für CBR-Systeme (3): QBIC

- **„Query By Image Content“, IBM**
  - Einführung der Mahalanobis-Distanz für Histogrammvergleich
  - Farbsuche im Lab-Raum, Textur über Coarseness, Contrast, Directionality
  - Formsuche: interaktive Segmentierung, Flächenmomente
  - Beschleunigung der Suche durch Dimensionsreduktion mittels Hauptkomponentenanalyse
  - Anfrage durch Beispiel oder Skizze
  - Integriert in DB/2 ImageExtenders
  - z.B. Suchmaschine der Ermitage in St. Petersburg

Beispiele für CBR-Systeme (4): QBIC

Beispielsuche Ermitage





1) A Game of Bowls

Matisse, Henri 1908

3) A Corner in Turin

Premazzi, Luigi
Mid–19th century

5) Cottages

Gogh, Vincent van
1890

Beispiele für CBR-Systeme (5): PhotoBook

- **Zusammenfassung von mehreren CBR-Projekten am MIT**
  - Einführung der *eigenface*-Methode zur Gesichtserkennung
  - Multikanalmethoden zur Textursuche - WOLD-Merkmale:
    - spektrale Zusammensetzung - Periodizität, Orierntierung
    - Farbe
    - Parameter eines autoregressiven Modells
  - *Modal Matching* für Formanalyse
  - Interaktive Segmentierung, halbautomatische Klassifikation
  - große Datenbanken zur Validierung (Textur, Gesichter)

Beispiele für CBR-Systeme (6): Virage Image Search Engine

- **Virage: erste erfolgreiche start-up Firma für CBR**
- ***framework* für die Entwicklung von CBR-Anwendungen:**
  - Merkmale: globale und locale Farbe, Textur, Form, erweiterbar
  - Ähnlichkeitssuche: Histogramme, Kombination, erweiterbar
  - plug-ins für Informix und Oracle
  - Grundlage des AltaVista Photofinders
- **Neuerdings: Konzentration auf Video:**
  - technische Fragen: Speicherung und *streaming*
  - *shot detection/keyframe extraction*, visuelle Indexierung
  - parallele Erkennung von Text im Bild, Gesichtern, Audio (Sprechererkennung, Typen)

Beispiele für CBR-Systeme (7): Aktivitäten in Deutschland

- **verschiedene Forschungsgruppen, aber relativ klein, viele haben wieder aufgegeben**
  - U Bremen (Herzog): PictureFinder
  - U München (Seidl): Ähnlichkeitssuche
  - U Freiburg (Burkhardt): Noblesse (EU-Projekt)
  - U Mannheim (Effelsberg). Automatic Video Abstraction
  - FhG Institut Graphische Datenverarbeitung (Volmer): FIR
- **Firmen:**
  - tecmath (Kaiserslautern): Bildarchive für Medienindustrie (eigene Entwicklung, jetzt hauptsächlich Schlüsselwortsuche)
  - Cobion (Kassel): Suche im Internet (Gesichter, Marken, digitale Wasserzeichen, eigene Entwicklung?)