

IV.2 Content-based Image Retrieval

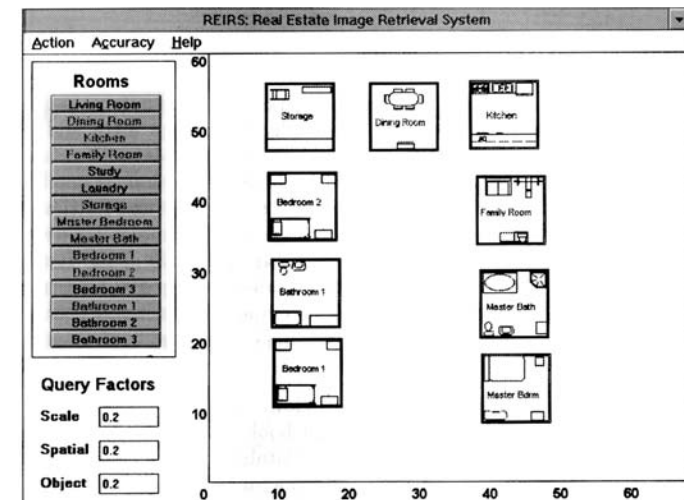
A Taxonomy of Image Retrieval Types

- ❑ *retrieval by browsing (RBR)*—browsing may be performed on actual images or “thumbnails”
- ❑ *retrieval by objective attributes (ROA)*—a query is formulated using meta attributes, logical attributes or a combination (similar to database queries—perfect match)
- ❑ *retrieval by spatial constraints (RSC)*—queries are based on relative spatial relationships among image objects, e.g. directional relationships, adjacency, overlap and containment
 - ❑ *relaxed RSC queries*—retrieve images that satisfy as many desired spatial relationships as possible (ranking required)
 - ❑ *strict RSC queries*—retrieve images that precisely satisfy all the spatial relationships specified in the query image

A Taxonomy of Image Retrieval Types

- ❑ *retrieval by spatial constraints (con't.)*
 - ❑ few objects involved in a query—explicit specification of desired spatial relationships possible
 - ❑ many objects involved in a query—a query is specified by placing icons (corresponding to domain objects) in a special window (*sketch pad window, spw*)
- The sketch pad window provides both the graphic icons and the necessary tools for selecting and placing these icons for composing an RSC query. Spatial relationships among icons implicitly indicate the desired SC among objects in images to be retrieved.

Sketch Pad Window



A Taxonomy of Image Retrieval Types

- ☐ *retrieval by semantic attributes (RSA)*—a query is specified in terms of the domain concepts
- ☐ *retrieval by feature similarity (RFS)*—a query is specified by selecting features of example images; images are retrieved that contain objects with similar features; often referred to as “content-based” retrieval

Motivation

- ☐ Konventionelle Formen
 - ☐ Formale Erfassung + Inhaltliche Erschließung
 - ☐ Inhaltliche Erschließung enorm aufwändig
 - ☐ manchmal nicht durchführbar (Bilder im Internet)
- ☐ Content Based Image Retrieval
 - ☐ Bilder aufgrund des dargestellten Inhalts finden
 - ☐ Merkmale aus Bildern extrahieren
 - ☐ Ähnlichkeitsmessung (keine exakte Übereinstimmung der Bilder)

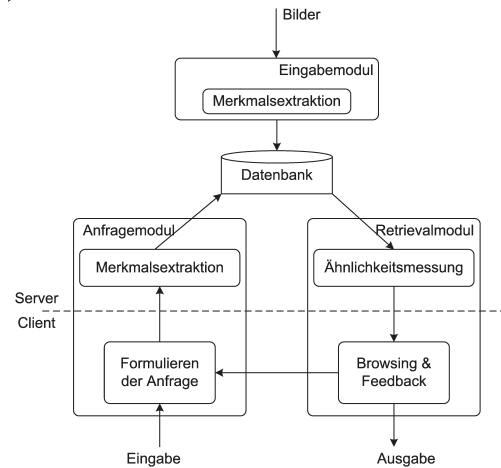
Anwendungsgebiete

- ☐ Medienagenturen
 - ☐ große Bilddatenbanken
 - ☐ manuelle Vergabe von Schlüsselwörtern ist zeit- und kostenintensiv
- ☐ Markenzeichen
 - ☐ Prüfung, ob ähnliche Markenzeichen existieren
- ☐ Produktkataloge
 - ☐ Kunde erinnert sich eher an Aussehen eines Produkts als an dessen Name oder Artikelnummer

Anwendungsgebiete

- ☐ Kriminalistik
 - ☐ Verfahren zur Gesichtserkennung
 - ☐ Identifikation von Fingerabdrücken
- ☐ Medizin
 - ☐ Suche ähnlich gelagerter Krankheitsfälle
 - ☐ Vergleich von Krankheitsbildern

CBIR-Systemarchitektur



Anfragearten

- ❑ Browsing
 - ❑ Benutzer interagiert mit dem System
 - ❑ stöbern in vorgegebenen Kategorien
- ❑ Textsuche
 - ❑ Suche mit Schlagwörtern
- ❑ Visuelle Beschreibung
 - ❑ query-by-example – Vorgabe eines Beispielbildes
 - ❑ query-by-sketch – Benutzer skizziert gesuchtes Bild (abhängig von Talent, Technik, Genauigkeit, ...)
 - ❑ query-by-template – Benutzer definiert Suchbild durch Auswahl aus Farb- und Texturkomponenten

IV.2.1 Merkmale

Merkmale – Überblick

- ❑ Primitive Merkmale
 - ❑ Auch Wahrnehmungsmerkmale und Low-Level-Features genannt
 - ❑ Einfache Attribute wie Farbe, Textur, Form, räumliche Beziehung von Objekten
 - ❑ Werden automatisch aus Bildern extrahiert und miteinander verglichen

Merkmale – Überblick

- ☐ Semantische Merkmale
 - ☐ Erkennen von Objekten
 - ☐ System benötigt Wissen über Aussehen der Objekte
 - ☐ Rollen (Tennisspieler) und Szenen (Sonnenuntergang)
 - ☐ Problematik liegt im Aufstellen des Modells (Größe, Form, Bewegung, ...)

2.1.1 Farbmerkmale

PRODUCTS & TECHNOLOGIES
VIR
 TECHNOLOGY DEMO



CLICK AN IMAGE TO FIND SIMILAR IMAGES

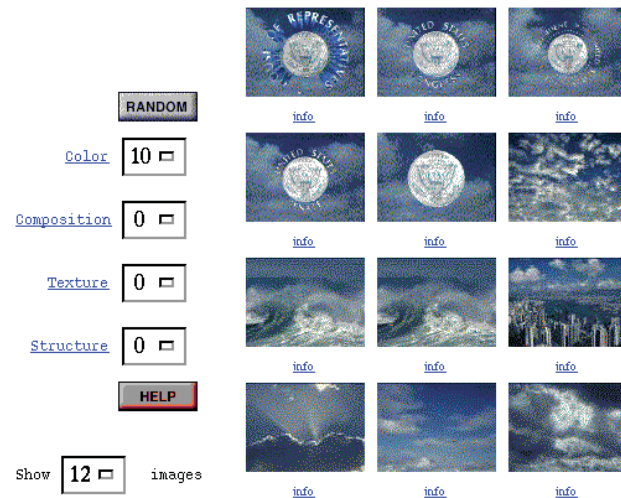


Image Features—Color

The Global Property of Color Distribution

- ☐ color histograms: histogram with e.g., 64 colors (bright red, pink, orange, ..., light blue, dark blue)

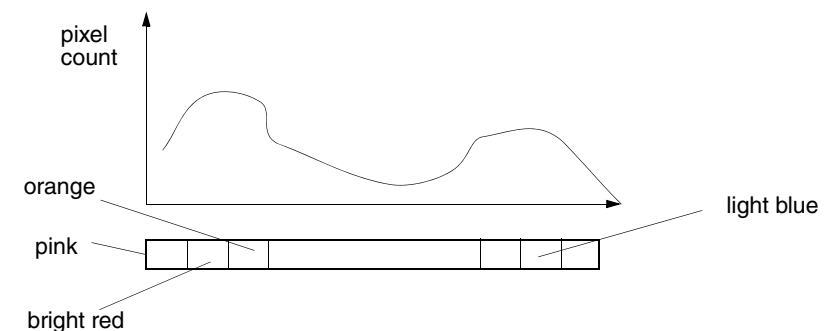


Image Features—Color

The Local Property of Object Color

- ❑ object detection by color-based segmentation—grouping of adjacent pixels with similar color properties into common regions; performance of the approach is fairly good when applied to images with reasonably homogeneous regions, but may be degraded in the presence of strong highlights or shading.

Farbmodelle

- ❑ RGB
 - ❑ Gleichmäßige Quantisierung führt zu redundanten und fehlenden Farbwerten
 - ❑ Distanzmaße nicht zufriedenstellend (z.B. Euklidische Distanz)
- ❑ Hue-Saturation-Brightness
 - ❑ Sowohl uniforme (z.B. $L^*u^*v^*$) als auch nicht-uniforme Farbräume (z.B. HSV)
 - ❑ Bei uniformen Farbräumen Ähnlichkeitsmessung durch Euklidische Distanz möglich

Histogramm – Eigenschaften

- ❑ Einfache und effiziente Repräsentation des Farbinhalts
- ❑ Einfache Berechnung
- ❑ Unempfindlich bezüglich Änderung der Bildauflösung, Kamerarotation und Zoomen
- ❑ Empfindlich gegenüber Änderung der Lichtverhältnisse
 - ❑ Verbesserung bei RGB durch Normalisieren der Farbwerte
 - ❑ Verbesserung bei HSB durch nicht Berücksichtigen der Helligkeitswerte
- ❑ Keine Information über die räumliche Verteilung der Farbe

Histogramm – Distanzmaße

Distanz zwischen Histogrammen – Maß für Ähnlichkeit

- ❑ L1-und L2-Maß (Euklidische Distanz)

$$D_H(I_Q, I_D) = \sum_{j=1}^n |H(I_Q, j) - H(I_D, j)|$$

$$D_H(I_Q, I_D) = \sqrt{\sum_{j=1}^n (H(I_Q, j) - H(I_D, j))^2}$$

Histogramm – Distanzmaße

☐ Histogram intersection

- ☐ unabhängig von der Anzahl der Bins im Histogramm

$$D_H(I_Q, I_D) = \frac{\sum_{j=1}^n \min(H(I_Q, j), H(I_D, j))}{\sum_{j=1}^n H(I_D, j)}$$

☐ Weighted Euclidean Distance

- ☐ berücksichtigt Zusammenhang zwischen einzelnen Bins

$$D_H(I_Q, I_D) = (H(I_Q) - H(I_D))^t A (H(I_Q) - H(I_D))$$

MPEG-7 Farbdeskriptoren

Überblick:

- ☐ Dominant Color – vorherrschende Farben repräsentieren Bilder bzw. Regionen
- ☐ Scaleable Color – ein mittels Haar-Transformation kodierte Farbhistogramm
- ☐ Color Structure – Farbhistogramm kombiniert mit räumlicher Information
- ☐ Color Layout – räumliche Anordnung der Farben im Bild
- ☐ Group of Pictures – Erweiterung von Scaleable Color auf eine Bildfolge eines Videos

MPEG-7 Farbdeskriptoren

Überblick:

zur zusätzlichen Beschreibung der obigen Deskriptoren

☐ Color Space

- ☐ unterstützt werden RGB, YCbCr, HSV und HMMD

☐ Color Quantization

- ☐ uniforme Quantisierung eines Farbraums (Definition der Bins)
- ☐ wird üblicherweise mit Dominant Color Deskriptor verwendet

MPEG-7 Farbdeskriptoren

☐ Dominant Color

- ☐ N dominante Farben plus Prozentsatz der Pixel, die diesen Farben entsprechen
- ☐ Eigenschaften bzgl. räumlicher Verteilung
- ☐ kompakte und effiziente Beschreibung

MPEG-7 Farbdeskriptoren

- ☐ Scalable Color
 - ☐ Farbhistogramm im HSV-Farbraum
 - ☐ Quantisierung uniform, 256 Bins fix
 - ☐ kodiert mittels Haar-Transformation
 - ☐ skalierbar durch Weglassen von Koeffizienten
 - ☐ Ähnlichkeitsmessung durch Distanzberechnung zwischen Histogrammen (z.B. L1-Maß)

MPEG-7 Farbdeskriptoren

- ☐ Color Structure
 - ☐ Histogramm wird mit räumlicher Information erweitert
 - ☐ Quadrat von 8x8 Pixel über das Bild bewegt
 - ☐ enthaltene Farben tragen zum Histogramm bei
 - ☐ Farben im HMMD-Farbraum, nicht-uniforme Quantisierung
 - ☐ 32 - 256 Bins

MPEG-7 Farbdeskriptoren

- ☐ Color Layout
 - ☐ beschreibt räumliche Verteilung der Farben
 - ☐ 8x8-großes Feld gefüllt mit repräsentativen Farben der entsprechenden Bildregionen (Farbmittelwerte)
 - ☐ Farben im YCrCb-Raum, DCT
 - ☐ Auswahl und Kodierung von üblicherweise 12 Koeffizienten
 - ☐ Ähnlichkeitsmessung durch Vergleich der Koeffizienten
 - ☐ einfache und schnelle Berechnung

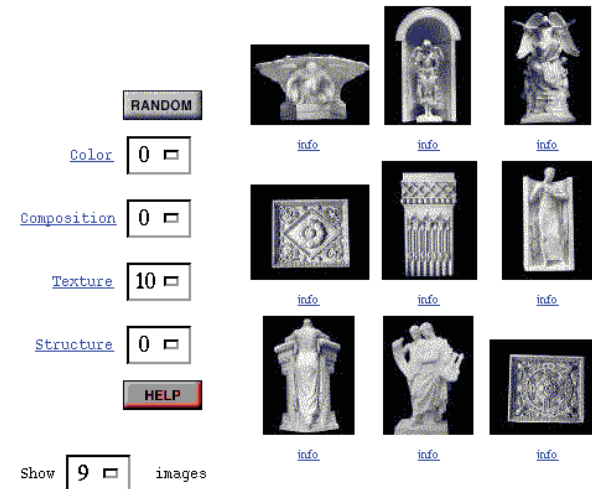
MPEG-7 Farbdeskriptoren

- ☐ Group of Frames/Group of Pictures
 - ☐ Erweiterung des Scaleable Color Deskriptor
 - ☐ Errechnung eines Histogramms aus den Farbhistogrammen der Einzelbilder
 - ☐ Mittelwert, Median oder Durchschnitt

2.1.2 Texturemerkmale



CLICK AN IMAGE TO FIND SIMILAR IMAGES



Überblick

- ☐ Texturen nur schwer beschreibbar (retrieval-by-example)
- ☐ Textur auch bei Graustufenbilder aussagekräftig
- ☐ Eigenschaften von Texturen
 - ☐ Körnigkeit – Feinheitsgrad der Textur
 - ☐ Periodizität – Regelmäßigkeit der Textur
 - ☐ Ausrichtung der räumlichen Anordnung der Textur
 - ☐ Eigenschaften sind abhängig von Entfernung zu Kamera und vom Betrachtungswinkel

Verfahren

- ☐ Strukturell – Lage und Ausrichtung der Strukturelemente
 - ☐ besonders geeignet für regelmäßige Texturen
- ☐ Statistisch – Verteilung der Helligkeitswerte im Bild
 - ☐ Autokorrelationsfunktion – Maß für die Körnigkeit
 - ☐ Co-Occurence-Matrix – Kontrast
 - ☐ Simultaneous Auto Regressive Model (SAR) – Pixel werden durch die Helligkeitswerte ihrer Nachbarpixel ausgedrückt (RISAR, MRSAR)

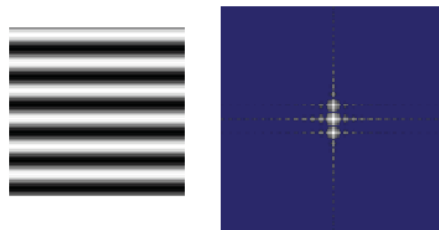
Verfahren

- ❑ Markov Random Fields
 - ❑ stochastisches Modell – Pixel werden in Abhängigkeit ihrer Nachbapixel modelliert
- ❑ Fraktale Modelle
 - ❑ zur Beschreibung von sehr unregelmäßigen Texturen
 - ❑ Beispiel – Erosionen, Küstenlinien

Texturanalyse im Frequenzbereich

- ❑ Frequenzbereich
 - ❑ beschreibt räumliche Verteilung der Helligkeitswerte
 - ❑ häufige Helligkeitswechsel entsprechen hohen Frequenzen
 - ❑ errechnet durch Fourier-Transformation
- ❑ Frequenzbild
 - ❑ Position gibt die Frequenz an
 - ❑ tiefe Frequenzen im Mittelpunkt
 - ❑ je weiter entfernt vom Mittelpunkt desto höher die Frequenz

Textur und Frequenzbild



Texture — Tamura Model 1

Three features:

- ❑ Contrast—statistical distribution of pixel intensities:

$$F_{con} = \frac{\sigma}{\alpha_4^{1/4}}$$

$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$

σ ...standard deviation, μ_4 ...fourth moment

- ❑ Coarseness — F_{crs} ... measure of granularity; is based on moving averages over windows of different sizes for each pixel

Texture — Tamura Model 2

- Directionality — gradient vector at every pixel magnitude (gray-level difference) and angle:

$$|\Delta G| = (|\Delta_H| + |\Delta_V|)/2$$

$$\theta = \arctan(\Delta_V/\Delta_H) + \frac{\pi}{2}$$

A histogram of θ values corresponding to magnitudes larger than a given threshold is constructed:

- highly directional images — histogram with strong peaks
- non directional images — flat histogram

An overall directionality measure F_{dir} based on the sharpness of the peaks

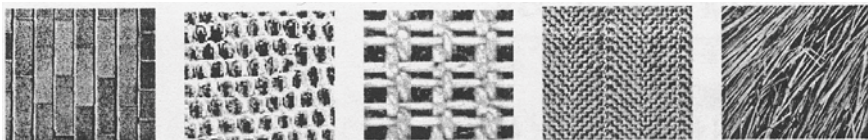
Texture — Other Models

- RISAR — Rotation Invariant SAR
- MRSAR Model — MultiResolution SAR; Describing granularities by representing textures at a variety of resolutions → SAR or RISAR applied to each level
- Wold features — texture is described by
 - periodicity—high, for strong harmonic components
 - directionality—high, for a strong evanescent component
 - randomness—high, for less structured texture

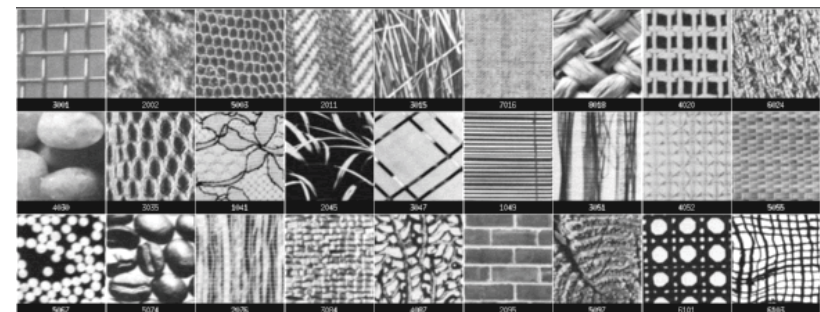
Texture — Brodatz Database

Brodatz is the *de facto* standard for evaluating texture algorithms

- Brodatz database— Photo album, which contains 112 gray-level images (512x512 8-bit), each represents a different texture class.
- Some Examples:



Brodatz Database — Further Examples



Multimedia

IV.2.1 Merkmale

Texture — Brodatz Database

Evaluating Texture Algorithm by Brodatz

9 sub-images (128x128) are extracted from the center of each Brodatz-image. Features are computed by applying the model to be evaluated to each of these 1008 sub-images. The distance of each test sub-image to all the others is then computed.

Two Results:

- retrieval rate — percentage of the n nearest subimages, belonging to the same class as the test sub-image
- average retrieval rate — each sub-image is computed as a test sub-image

428Object ColorChristian Breiteneder

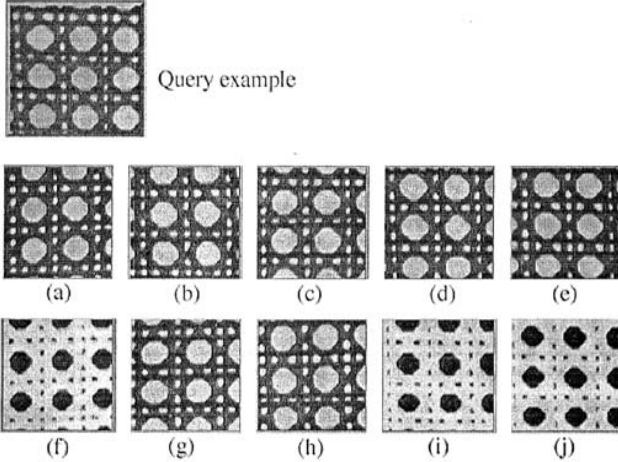
Multimedia

IV.2.1 Merkmale

Texture — Brodatz Database

Brodatz — Tamura Features

Example — Evaluating Tamura texture algorithm by Brodatz



429Object ColorChristian Breiteneder

Multimedia

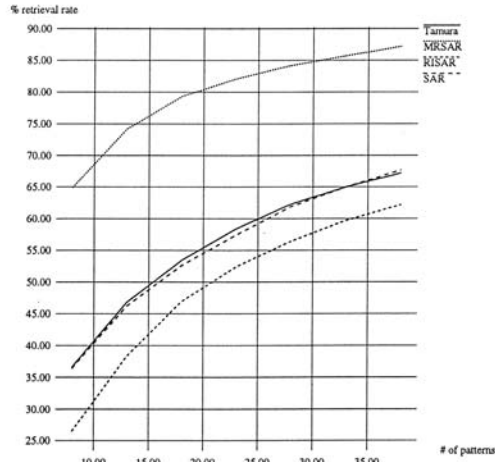
IV.2.1 Merkmale

Texture — Brodatz Database

Brodatz — Comparison

Average retrieval values for different values of n

8, 13, 18, 23, 28, 33, 38



430Object ColorChristian Breiteneder

Multimedia

IV.2.1 Merkmale

Brodatz — Comparison

Brodatz — Comparison

Texture algorithm comparison — time required for feature computation for a 128 x 128 sub-image on Sun SPARCstation 10 platform:

Algorithm	Time (sec)
SAR	1.6
RISAR	2.1
MRSAR	2.3
Tamura	9.9

Retrieval can be accomplished in real-time when image features are pre-computed.

431Object ColorChristian Breiteneder

Brodatz — Comparison

❑ Discussion —

Tamura

- ❑ disadvantages - same accuracy but slower than SAR
- ❑ advantages - rotation invariant; Tamura features closer to human visual perception

MRSAR

- ❑ advantages - best accuracy
- ❑ disadvantages - computing lower resolutions of the image when the image region is too small to start with is not possible

MPEG-7 Texturdeskriptoren

Überblick:

- ❑ Texture Browsing Descriptor
 - ❑ beschreibt Ausrichtung, Regelmäßigkeit und Körnigkeit der Textur
- ❑ Homogeneous Texture Descriptor
 - ❑ Beschreibung regelmäßiger Texturen
- ❑ Edge Histogram Descriptor
 - ❑ räumliche Verteilung von Kanten im Bild

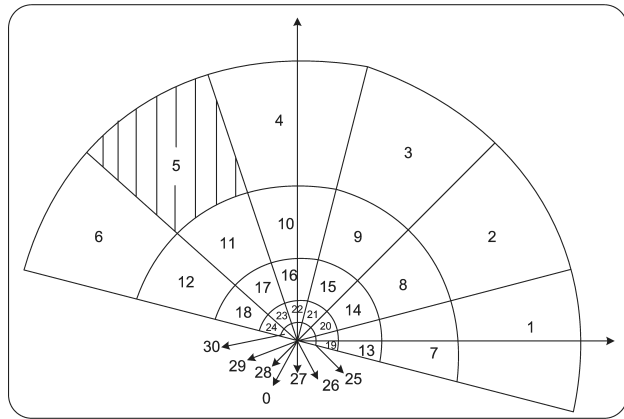
Texture Browsing Descriptor

- ❑ Ausrichtung (direction)
 - ❑ beschreibt vorherrschende Richtung der Textur
 - ❑ 6 verschiedene Richtungen (0° bis 150° in 30° - Schritten)
 - ❑ 2. Richtung optional
- ❑ Körnigkeit (coarseness)
 - ❑ Feinheitsgrad für jede Richtung der Textur – 0 (fein) bis 3 (grob)
- ❑ Regelmäßigkeit (regularity)
 - ❑ 4 Abstufungen für die Regelmäßigkeit – 0 (gänzlich unregelmäßig) bis 3 (gleichmäßiges Muster)

Homogeneous Texture Descriptor

- ❑ Beschreibung regelmäßiger Texturen
 - ❑ wichtig für Luft- und Satelliten-Aufnahmen
- ❑ Berechnung basiert auf Frequenzbild

Homogeneous Texture Descriptor



2.1.3 Formmerkmale

Edge Histogram Descriptor

- ☐ beschreibt räumliche Anordnung von Kanten im Bild
- ☐ auch für sehr unregelmäßige Muster geeignet
- ☐ häufig kombiniert mit Farbdeskriptor (z.B.: ColorLayout Descriptor)
- ☐ 5 Arten von Kanten
 - ☐ 4 richtungsbezogene (0° , 45° , 90° und 135°)
 - ☐ 1 nichtrichtungsspezifische (isotrop)

Formmerkmale

- ☐ 2 Arten von Formen
 - ☐ umrissbasiert (contour shape)
 - ☐ Form eines Objekts aufgrund seiner Umrisslinie repräsentiert (Kantenextraktion)
 - ☐ bereichsbasiert (region shape)
 - ☐ zusammenhängende oder unterbrochene Regionen

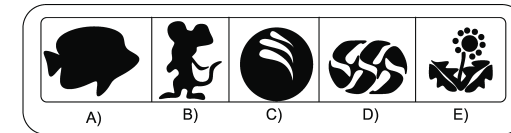
MPEG-7 Shape Descriptors

Vier Methoden zur Beschreibung von Formeigenschaften

- ☐ Region Shape – Pixel innerhalb eines bestimmten Bereichs (Objektmomente)
- ☐ Contour Shape – Umriss eines Objekts (Curvature Scale Space Darstellung)
- ☐ Shape 3D Spectrum – Anzahl von lokalen Konvexitäten einer 3D Oberfläche (Histogrammform)
- ☐ Shape 2D/3D – Beschreibung einer 3D Form durch mehrere 2D Deskriptoren (aus verschiedenen Winkel)

Region Shape Descriptor

Region-Shape-Descriptor beschreibt sowohl einfache (a, b) als auch komplexe (c, d, e) Formen



Region Shape

- ☐ Transformationsinvariante Objektmomente
- ☐ Der Region-Shape-Descriptor beschreibt
 - ☐ einfache Formen
 - ☐ komplexe Formen
- ☐ Verfahren zur Berechnung von Region Shape Deskriptoren
 - ☐ GMD – Geometric Moment Descriptor
 - ☐ GD – Grid Descriptor
 - ☐ ZMD – Zernike Moment Descriptor

Contour Shape

- ☐ Beschreibung Umriss eines Objekts / Objektausschnittes
- ☐ Curvature Scale Space (CSS)
- ☐ Vorgehensweise
 - ☐ Berechnung der Randkoordinaten

$$(x(t), y(t)), t = 1, 2, \dots, N - 1$$
 - ☐ Normieren, sodass Objekte mit unterschiedlicher Anzahl von Randkoordinaten verglichen werden können
 - ☐ Umriss des Objekts wird geglättet bis keine Wendestellen mehr existieren

Contour Shape

❑ Vorgangsweise (2)

❑ Glättung mittels Gauß-Filter:

$$x'(t) = x(t) * g(t, \sigma), y'(t) = y(t) * g(t, \sigma)$$

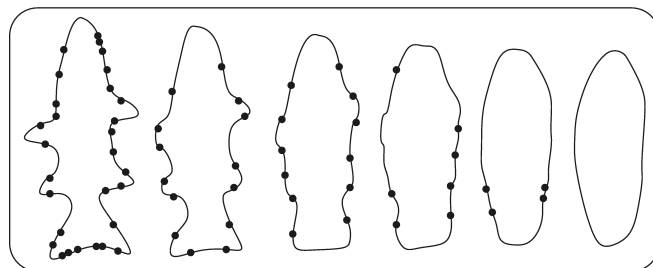
- ❑ Bestimmung von Wendestellen relativ zu einem fixen Startpunkt zu jedem Glättungsschritt
- ❑ Erstellen des CSS-Diagramms
- ❑ zu jeder Konvexität oder Konkavität der Kontur existiert ein Peak im CSS-Bild
- ❑ Ermittlung der Lage und Höhe der Maxima der einzelnen Peaks

Contour Shape

❑ Vorgangsweise (3)

- ❑ Ähnlichkeitsmessung durch Histogrammvergleich
- ❑ invariant gegenüber Skalierung und Rotation

Glättung



Shape 3D Spectrum

- ❑ Beschreibung lokaler geometrischer Eigenschaften
- ❑ Shape Index (SI)
 - ❑ beschreibt Konvexität in einem Punkt
 - ❑ Funktion der beiden Hauptkrümmungen in diesem Punkt

$$S I_p = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{k_p^1 + k_p^2}{k_p^1 - k_p^2}$$

- ❑ skalierungs-, translations- und rotationsunabhängig
- ❑ elementare Formen

Shape 2D/3D

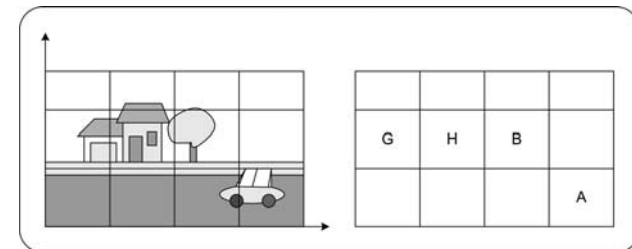
- ☐ 3D Objekt wird aus verschiedenen Winkeln als 2D Bild beschrieben
- ☐ Ähnlichkeitsvergleich mit Bildern aus gleichen Betrachtungswinkel
- ☐ Anwendung aller visuellen Deskriptoren
 - ☐ contour shape
 - ☐ region shape
 - ☐ color descriptors
 - ☐ texture descriptors

2.1.4 Weitere Merkmale

Weitere Merkmale

- ☐ Räumliche Anordnung von Objekten in einem Bild
- ☐ Kantenbilder (query-by-sketch)
- ☐ Beziehung von Objekten zueinander
 - ☐ Richtungsbezogen – Position zueinander (links, rechts etc), Entfernung, Winkel zwischen den Objekten, ...
 - ☐ Topologie – „gleich wie“, „innerhalb von“, „berührt“, „überlappt“, Beziehungen sind invariant gegenüber Translation, Rotation, Skalierung

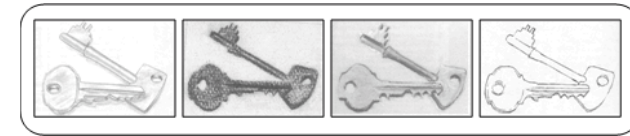
Beispiel – 2D-String



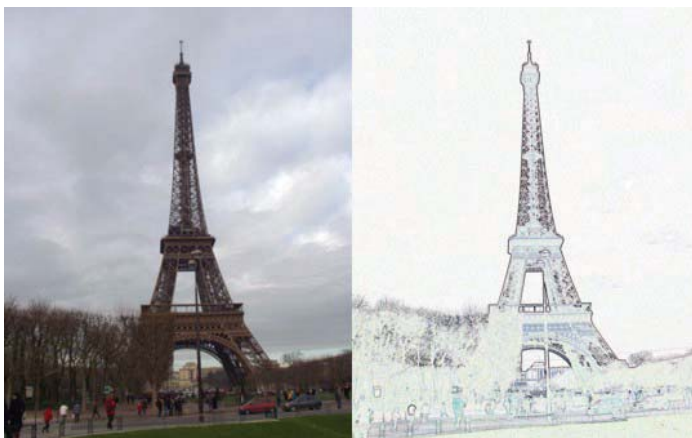
Beispiel – 2D-String

- ☐ 2D-String-Repräsentation
 - ☐ Projektion der Objekte auf die x-Achse – h: $G < H < B < A$
 - ☐ Projektion der Objekte auf die y-Achse – v: $A < G = H = B$
- ☐ Erweiterungen zu 2D-String
 - ☐ Berücksichtigung von Entfernung und Ausrichtung der Objekte
 - ☐ zusätzliche Beziehungen (A überlappt B, A enthält B, ...)
 - ☐ Umgang mit Segmenten von Objekten
- ☐ 2D-String Repräsentation ändert sich bei Translation nicht

Kantenbilder – Skizze



Kantenbilder – Original und Kantenbild



Kantenbilder – Query-by-Sketch

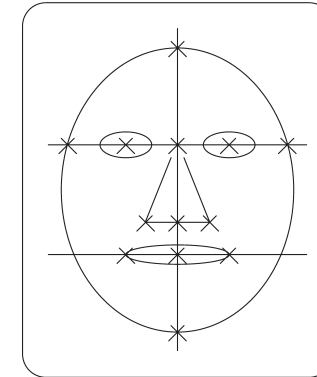
- ☐ query-by-sketch
 - ☐ Bildsuche mittels handgefertigter Skizze
 - ☐ Erfolgsquote abhängig von Fähigkeit des Zeichners und Genauigkeit der Skizze
 - ☐ Merkmale abhängig von verwendeter Technik
 - ☐ Umriss, Schattierung, Farben
- ☐ Kantenbilder
 - ☐ Bilder müssen als Kantenbilder repräsentiert werden
 - ☐ Position der Objekte muss nicht exakt stimmen

Gesichtserkennung

- ☐ 2 Aufgabenstellungen
 - ☐ Gesichtsdetektion – das Erkennen von Gesichtern in Bildern
 - ☐ Gesichtsidentifikation – das Wiedererkennen eines Gesichts
- ☐ Anwendungsgebiete
 - ☐ Zugangskontrollen für sicherheitskritische Bereiche
 - ☐ Identifizierung von Straftätern

Gesichtserkennung

Charakteristische Punkte von Gesichtern



Gesichtserkennung

- ☐ Geometrische Merkmale
 - ☐ Abstand der Augen, Nasenlänge, Breite des Mundes, ...
 - ☐ horizontale und vertikale Kanten des Gesichts ermitteln
 - ☐ charakteristische Punkte des Gesichts
- ☐ Ähnlichkeitsmessung
 - ☐ Distanzen einzelner Punkte zueinander – abhängig von Lage, Ausrichtung und Größe des Gesichts
 - ☐ Normierung der Distanzen
 - ☐ Relationen von Distanzen – Abstand rechtes Auge zum rechten Rand zu Abstand linkes Auge zum linken Rand

Verfahren zur Gesichtserkennung

- ☐ Template Matching
 - ☐ Erstellung von Schablonen für Gesichter bzw. Gesichtsteile (Mund, Nase, ...)
 - ☐ Maß der Ähnlichkeit – Übereinstimmungen der Schablonen zueinander
- ☐ Ähnlichkeitsmessung im Frequenzbereich
 - ☐ Diskrete Fouriertransformation
 - ☐ Maß der Ähnlichkeit wird durch Koeffizientenvergleich bestimmt

Verfahren zur Gesichtserkennung

- ❑ Eigenfaces
 - ❑ Bilder werden als Vektoren repräsentiert (Komponente = Pixel des Bildes)
 - ❑ Bilder, die Gesichter darstellen, bilden Unterraum des gesamten Vektorraums
 - ❑ Eigenvektoren des Unterraums mittels PCA ermitteln
 - ❑ Gesichter werden durch Linearkombination der Eigenvektoren approximiert (Eigenfaces)
 - ❑ Maß der Ähnlichkeit – Distanz zweier Eigenfaces

Probleme bei Gesichtserkennung

Faktoren, die die Gesichtserkennung erschweren:

- ❑ Bildqualität (Schärfe, Kontrast, Helligkeit)
- ❑ unterschiedliche Beleuchtung, Probleme mit harten Schatten
- ❑ unterschiedliche Kamerawinkel
- ❑ Kopfposition und -haltung (Frontalansicht, Seitenansicht)
- ❑ Gesicht unterschiedlich durch Bart, Make-Up, Tragen einer Brille, ...
- ❑ Gesichtsausdruck verändert das Gesicht (traurig, wütend, fröhlich, ...)

User Interface



Figure 11.13 A menu of sample images for query by visual templates.

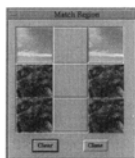


Figure 11.13 (continued)
A query image composed from selections from the menu.

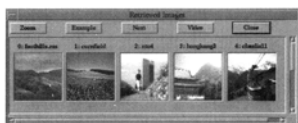


Figure 11.13 (continued)
The result of retrieval from the composed query.

- ❑ Query by visual templates — retrieve images containing some known color/texture patterns (sea, lawn, forest...):
Template map: 9 sub-areas
Different templates selected visually from their icons or from their names

User Interface For Formulating Queries 2

- ❑ Drawing a Query — user draws a sketch with a pen-based device or photomanipulation-tool



User Interface For Formulating Queries 3

- ❑ Query by visual examples — user may formulate an uncomplete query which retrieves a preliminary set of images. From these he selects one visually similar to the desired target. Query based on either the entire example image or a specific region of image.

Example: Blobworld content-based retrieval system of the University of Berkley (<http://elib.cs.berkeley.edu/photos/blobworld/start.html>)

Query in Blobworld

Blobworld —

- ❑ Features — color (*218 bin histogram*), texture (*mean contrast, anisotropy*), location, shape of region=blobs (*area, eccentricity, orientation*)
- ❑ Query — user selects category; in a initial image user selects a blob, indicates the importance of blob's features. Keywords optional.
- ❑ Example — Query image selected from category people; keyword: portrait

Query in Blobworld 2

Query image: 388034 Query blob

Querying from 35000 images (136 have the keyword "portrait.")

feature importance:					
overall	color	texture	location	shape	
blob	very	very	very	very	very

5: 250093 (score = 0.73) New query

6: 53041 (score = 0.73) New query

7: 181040 (score = 0.7) New query

8: 358008 (score = 0.69) New query

1: 401024 (score = 0.81) New query

2: 357036 (score = 0.81) New query

3: 401038 (score = 0.79) New query

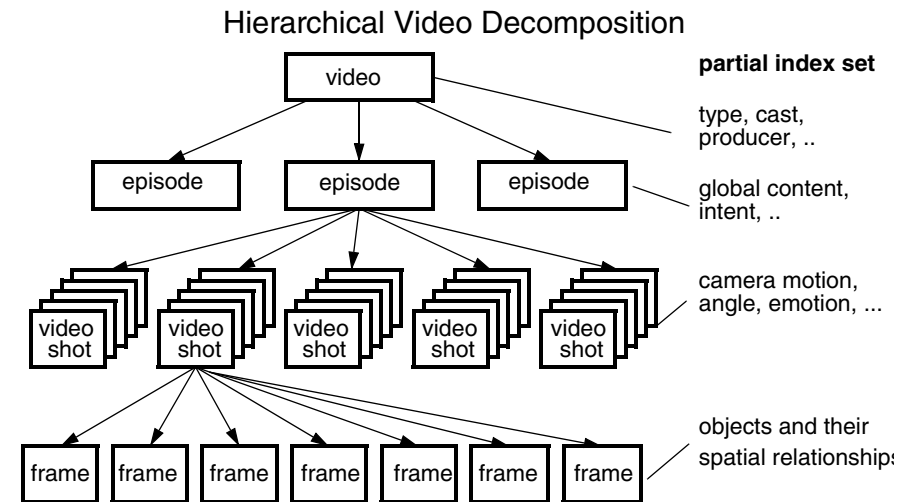
4: 401024 (score = 0.73) New query

Part IV Multimodal Information Retrieval

IV.3 Video Retrieval

3.1 Video Segmentation (Schnitterkennung)

Video Segmentation



Video Segmentation

Terminology:

- ❑ A *shot* or *take* refers to a contiguous recording of video frames depicting a continuous action in time and space.
- ❑ *Video segmentation* is the process of determining the boundaries between consecutive camera shots.

Shot-level organization of video is considered most appropriate for video browsing and content-based retrieval.

Video Segmentation

Work on video segmentation can be grouped into—

- ❑ frame difference techniques
 - ❑ pixel comparison
 - ❑ frame histogram comparison
- ❑ camera operation
- ❑ techniques on compressed video

Frame Difference Techniques

- ❑ **pair-wise pixel comparison**—a pixel is judged different if difference between intensity values in 2 successive frames exceeds a given threshold t

$$DP_i(k, l) = \begin{cases} 1 & \text{if } |P_i(k, l) - P_{i+1}(k, l)| > t \\ 0 & \text{else} \end{cases}$$

where (k, l) = the coordinates of pixels
 i = the index of the current frame

$DP_i(k, l)$ determines how many pixels have changed

Frame Difference (Example)

$$\text{abs}(P_i(k, l) - P_{i+1}(k, l)) = DP_i(k, l)$$



Frame Difference Techniques

- ❑ **segment boundary**—if more than a given percentage of the total number of pixels have changed

$$\frac{\sum_{k,l=1}^{M,N} DP_i(k,l)}{MN} \times 100\% > T_b$$

potential problem: metric is sensitive to camera motion, illumination changes, object movement and noise

Frame Difference Techniques

- ❑ **Likelihood ratio**—based on the *second-order statistical characteristics* of intensity values of corresponding *regions* (blocks) in successive frames.
- ❑ Two consecutive frames partitioned into a set of sample areas. Likelihood ratio for one sample area is computed by—

$$\frac{\left[\frac{S_i + S_{i+1}}{2} + \left(\frac{\mu_i - \mu_{i+1}}{2} \right)^2 \right]^2}{S_i S_{i+1}} > t$$

μthe mean intensity for a given region

Sthe variance for this region

iindex of current frame

Frame Difference Techniques

- ❑ A camera break can be declared if the total number of sample areas whose likelihood ratio exceeds the threshold is sufficiently large (depends on how the frame is partitioned)
- ❑ Advantage over *pair-wise pixel comparison*—likelihood ratio raises the level of tolerance to slow and small object motion from frame to frame. Makes it less likely that these effects will mistakenly be interpreted as camera breaks
- ❑ Disadvantage—if the two compared sample areas have same mean and variance, but completely different probability density function, no change will be detected. Fortunately, such a situation is very unlikely

Histogram Techniques

- ❑ histogram of a frame provides a global description; let $H_i(j)$ denote the histogram value for the i -th frame, where j is one of the G possible gray levels, the difference between i -th frame and its successor is given by

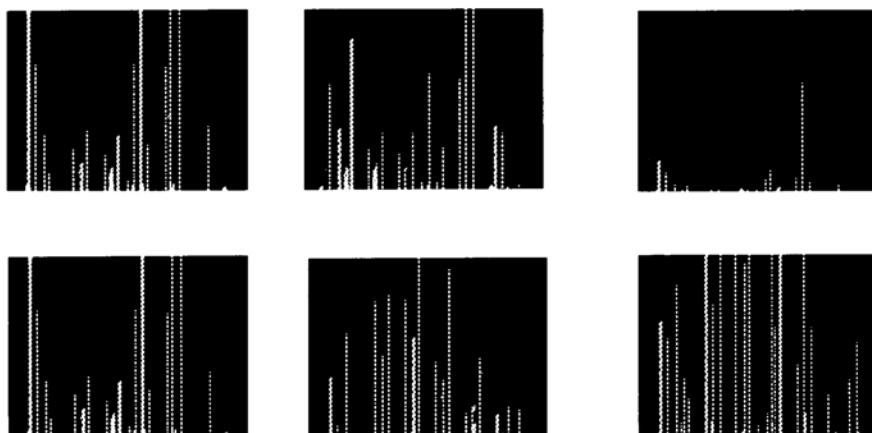
$$SD_i = \sum_{j=1}^G |H_i(j) - H_{i+1}(j)|$$

If the overall difference SD_i is larger than a given threshold, a segment boundary is declared

- ❑ equation can also be applied to color channel histograms
- ❑ histogram technique less sensitive to object motion than the pair-wisepixel comparison algorithm

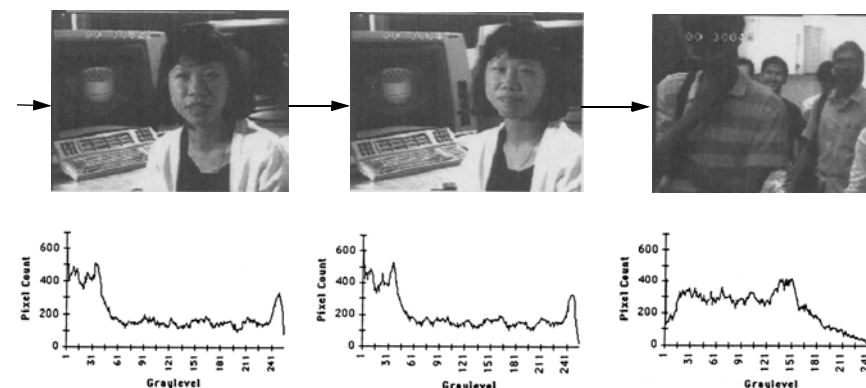
Frame-to-Frame Histogram Differences

$$H_i(j) \quad H_{i+1}(j) \quad \Rightarrow \quad \text{abs}(H_i(j) - H_{i+1}(j))$$

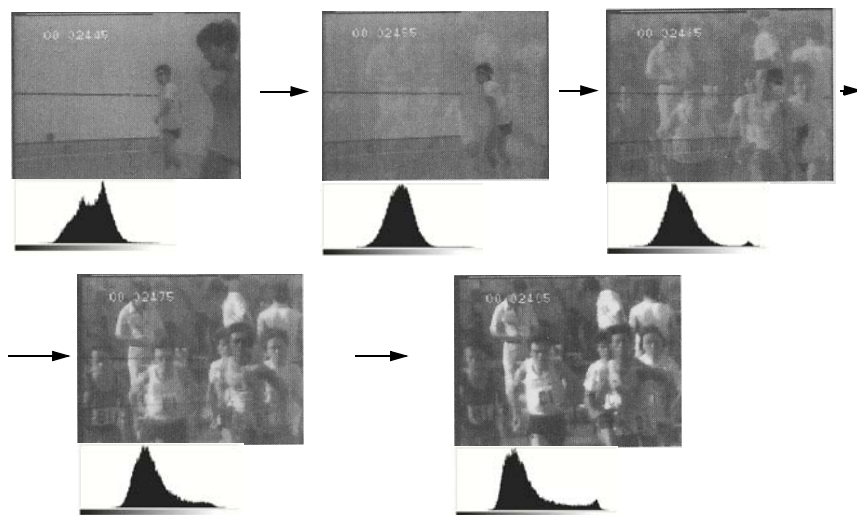


Histograms of a Camera Break

Three frames across a camera break and their histograms



Histograms of a Dissolve Sequence

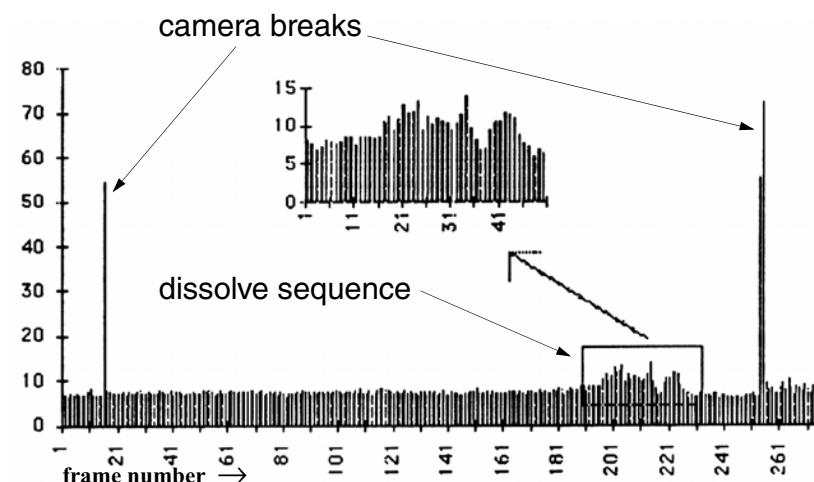


480

Video Segmentation

Christian Breiteneder

Frame-to-Frame Histogram Differences



481

Video Segmentation

Christian Breiteneder

Frame Histogram Differencing

The graph of frame-to-frame differences (previous page) exhibits 2 high pulses corresponding to camera breaks. However, the inset displays another sequence of pulses the values of which are higher than those of their neighbors, but lower than the threshold. This inset displays the difference values for a *dissolve sequence*.

What is required is a threshold value that will detect a dissolve sequence and distinguish it from a camera break. Similar approaches are applied to transitions implemented by other types of special effects.

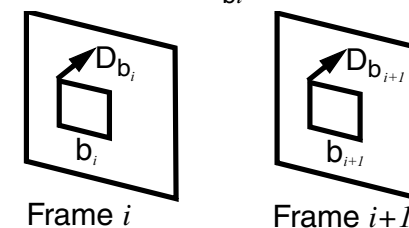
482

Video Segmentation

Christian Breiteneder

Motion Continuity

- ☐ Motion is an important property of any video source
- ☐ Optical flow—vectorfield, where each vector represents a velocity of the corresponding block on the image plane
- ☐ Motion vectors are computed by block matching, assigning to each block in frame i , b_i , a vector which displaces (i.e. translates) the entire block, D_{b_i}



483

Motion Continuity

Christian Breiteneder

Motion Continuity

Measure of motion smoothness—defined as the “ratio of velocity to motion”

- For each block b_i , $w_{1i}(b)$ is defined

$$w_{1i}(b) = \begin{cases} 1 & \text{if } |D_{bi}| > t_s \\ 0 & \text{otherwise} \end{cases}$$

- similarly, $w_{2i}(b)$ is defined

$$w_{2i}(b) = \begin{cases} 1 & \text{if } |D_{bi+1} - D_{bi}| > t_m \\ 0 & \text{otherwise} \end{cases}$$

- t_s and t_m small threshold values

Motion Continuity

- $\sum_b w_{1i}(b)$ —the number of significant motion vectors in frame i
- $\sum_b w_{2i}(b)$ —the number of motion vectors in frame i which differ significantly from their corresponding vectors in frame $i+1$
- the *smoothness* W_i of frame i —

$$W_i = \frac{\sum_b w_{1i}(b)}{\sum_b w_{2i}(b)}$$

if $W_i < T \Rightarrow$ camera break

Gradual Transitions

Consider the page “Frame-to-Frame Histogram Differences”

- Camera Break—high pulses
- Gradual Transitions—pulses with values higher than those of their neighbors but significantly lower than the cutoff threshold

Gradual Transitions-Examples

fade out



fade in



wipe



slide

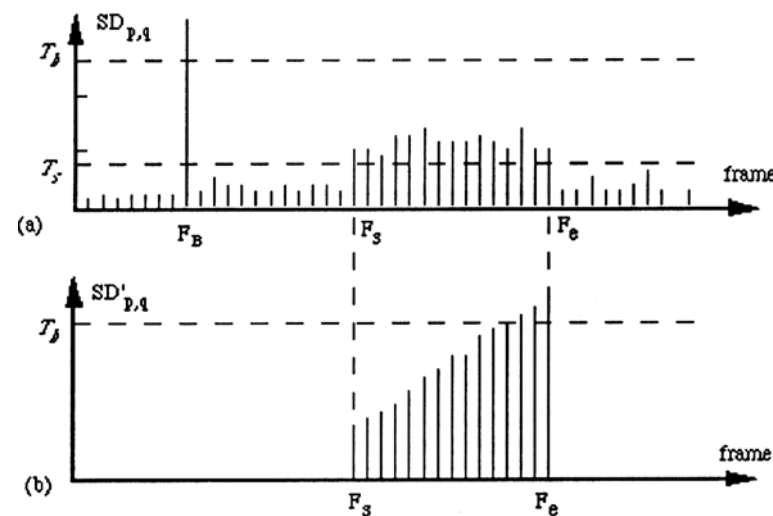


Twin-Comparison Approach

twin-comparison approach

- ❑ requires the use of two cutoff thresholds: T_b is used for camera break detection and T_s for special effect detection
- ❑ whenever a difference falls between T_s and T_b the corresponding frame is marked and compared to subsequent frames
- ❑ the end frame of the transition is detected when the difference between consecutive frames decreases to less than T_s , while the comparison with the frame marked has increased to a value larger than T_b .

Twin-Comparison Approach



Twin Comparison Approach

- ❑ $SD_{p,q}$ —the differences between consecutive frames defined by the difference metric
- ❑ $SD'_{p,q}$ —calculated when $T_b > SD_{p,q} > T_s$
the accumulated difference between the current frame and the potential starting F_s frame of a transition
- ❑ T_s —the threshold used to detect the starting frame F_s
- ❑ T_b —the threshold used to detect the ending frame F_e and camera breaks

Multi-Pass Approach

approach to reduce processing time

- ❑ First pass—
 - ❑ a lower T_b is set
 - ❑ resolution is sacrificed temporally (e.g. examining only 5frames/sec of a 25frames/sec source video) \Rightarrow large skip factor: increases $SD_{p,q}$ during gradual transitions
 - ❑ both camera breaks and gradual transitions will be detected (any number of false detections too)
- ❑ Further passes—increased temporally resolution is used to locate potential segment boundaries more accurately

3.2 Motion Analysis

Camera Operation Detection

- ❑ Changes associated with special effect transitions must be distinguished from changes introduced by camera movements.
- ❑ Changes due to camera movements tend to induce successive difference values of the same order as those of gradual transition. Therefore, it is necessary to detect patterns of image motion that are induced by camera operation
- ❑ Two approaches
 - ❑ motion vector analysis
 - ❑ (video X-ray)

Motion Vector Analysis

- ❑ frame i is partitioned into a number of blocks.
- ❑ *block matching* — each block is assigned a vector which displaces the entire block
- ❑ optical flow—vector field, where each vector represents a velocity of the corresponding block on the image plane
- ❑ the distribution of motion vectors in a frame may dedect
 - ❑ panning and tilting
 - ❑ zooming in and zooming out
 - ❑ particular motion patterns

Motion Analysis — Panning and Tilting

- ❑ Optical flow — predominatly the same direction, most of motion vectors parallel to a *modal vector*.

- ❑ Algorithm —

$$\text{if } \sum_{b=1}^N |\Theta_b - \Theta_m| \leq \Theta_p$$

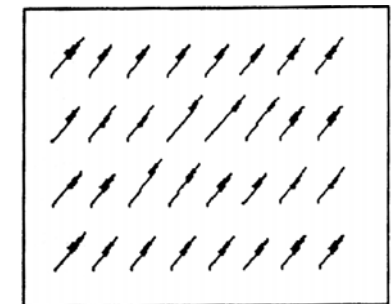
camera panning/tilting is defined

Ntotal number of blocks of the frame

Θ_mmodal angle value of the entire set

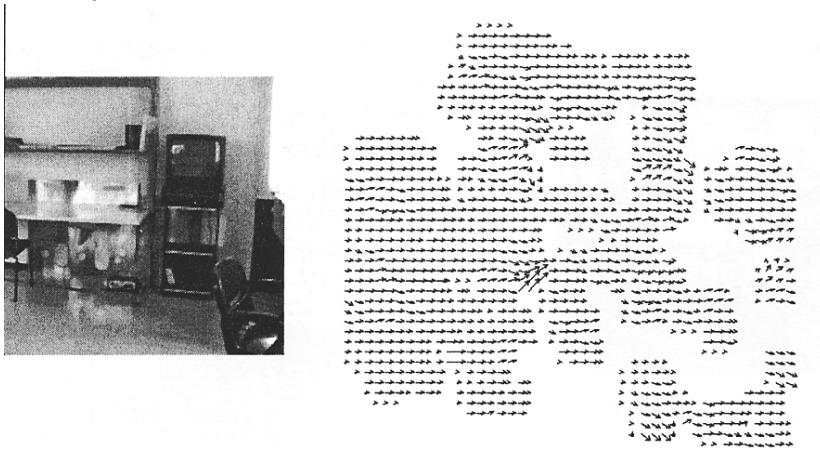
Θ_b direction (=angle) of the motion vector for block b

Θ_pthreshold value



Motion Analysis — Panning

Example—



496

Camera Operation Detection

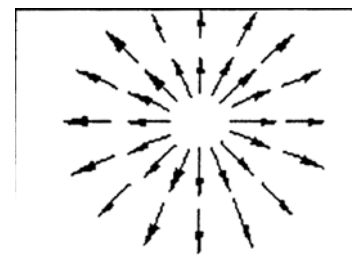
Christian Breiteneder

Motion Analysis — Zooming

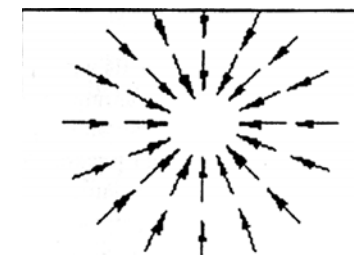
Assumption — focus in center of frame, no object movement

Algorithm—

if $\sum \text{motionvectors} = 0$ zooming is declared



pattern for zoom in



pattern for zoom out

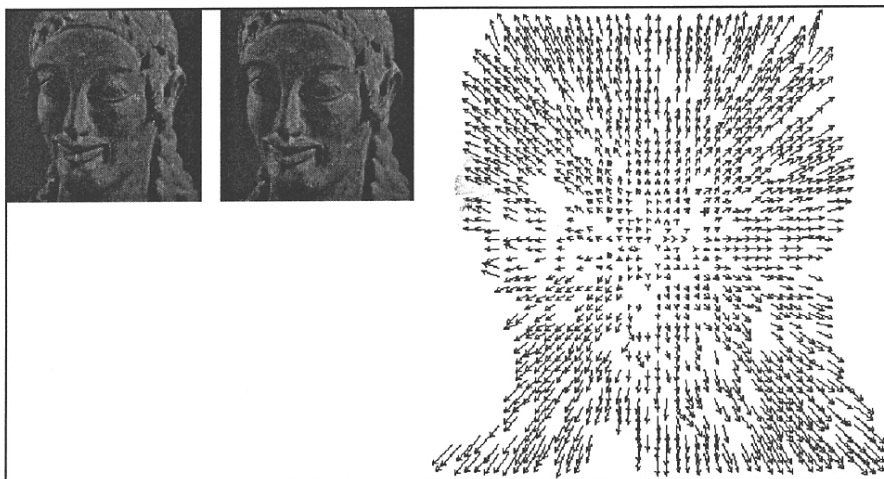
497

Camera Operation Detection

Christian Breiteneder

Motion Analysis — Zooming in

Example —



498

Camera Operation Detection

Christian Breiteneder

Motion Analysis — Zooming

Peripheral vision — this approach only requires the assumption that focus center is within the boundaries of a frame

$$\text{if } |v_k^{\text{top}} - v_k^{\text{bottom}}| \geq \max(|v_k^{\text{top}}|, |v_k^{\text{bottom}}|)$$

$$\text{and } |u_k^{\text{left}} - u_k^{\text{right}}| \geq \max(|u_k^{\text{left}}|, |u_k^{\text{right}}|)$$

zooming is declared

$v_k^{\text{top}}, v_k^{\text{bottom}}$ vertical components of motion vectors of the top and bottom row;

$u_k^{\text{left}}, u_k^{\text{right}}$ horizontal components of motion vectors for the left-most and right-most columns

499

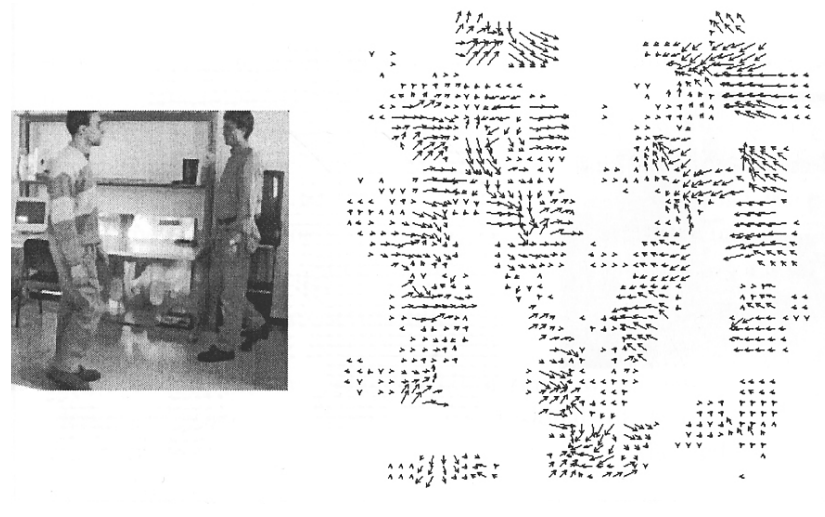
Camera Operation Detection

Christian Breiteneder

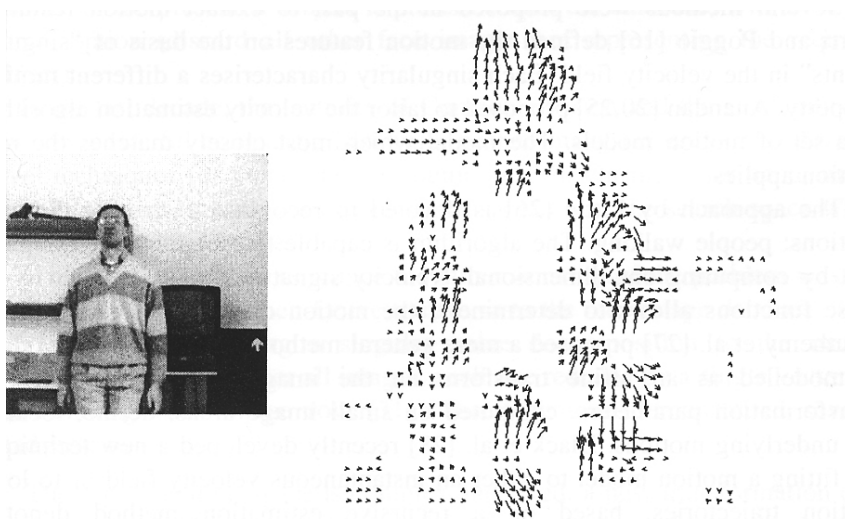
Motion Analysis — Object Motion

- ❑ Motion analysis is also used to define particular actions taking place in the scene—
 - ❑ human or animal motion — running, jumping, walking....
 - ❑ object motion — moving cars, trucks, airplanes, helicopters..
- ❑ A discription can be used to generate a rough discription of the scene—
 - ❑ Example— "an animal jumping viewed from a translating camera" is the rough discription of a video clip of a kangaroo taken from a car moving along with it.

Motion Analysis —Walking



Motion Analysis —Jumping



Camera Operation—Video X-Ray

Spatiotemporal image —

- ☐ assumption: background contains distinctive features
- ☐ panning—slanted lines in the top view
- ☐ tilting—slanted lines in the side view
- ☐ zooming—lines diverge or converge

