Kapitel 1 – Einführung

Multimedia: Multi = viele, Medium: Mittel zur Informationsverbreitung und Informationsdarstellung, die gleichzeitige Anwendung verschiedener Medien

Medien beeinflussen menschliche Computer-Interaktion:

- Sprechen ist schneller als schreiben.
- Hören ist leichter als lesen.
- Zeigen ist leichter als beschreiben.

Teilaspekte:

- Multimedialität: Präsentationsebene (Charakterisierung der Geräteausstattung/Mediums)
- Multicodalität: Codierungsebene (Nutzung verschiedener Symbolsysteme)
- Multimodalität: Perzeptionsebene (Ansprechen verschiedener Sinne)
  Virtual Reality Gesichtssinn, Gehörsinn, Taktile Darstellung, Geruch, Geschmack, Gleichgew.
- Interaktivität: Direkte Reaktion des Systems auf den Benutzer

Lokalisierungstechnologien:

- WLAN: Lokalisierung anhand Feldstärke/Signallaufzeit (ca. 5-10m Abweichung)
- Bluetooth: Lokalisierung anhand der Feldstärke, viele Landmarken notwendig (1m Abweich.)
- Ultra-Wide-Band: Lokalisierung anhand von Signallaufzeiten und Einfallswinkel (ca. 15cm Ab.)
- RFID: Lokalisierung anhand von Detektionsrate oder Feldstärke

Wiederholungen

Psychoakustisches Modell: Menschliches Ohr erfasst nur Klänge von 20 Hz bis 20k Hz. Erfassung nicht linear über gesamtes Frequenzspektrum und unscharf im Randbereich

Kapitel 2 – Bildverbesserung und Bildanalyse

Anwendungen: Zeichenerkennung, Qualitätsprüfung, Luftbildauswertung, Fahrzeugsteuerung, …

Bildverbesserung: Histogrammausgleich, Bildglättung, Bildverschärfung, Grauwerttransformation

Punktoperation: Ein Bildpunkt des Ausgabebildes ist nur eine Funktion eines einzelnen Bildpunktes

Lokale Operation: Ein Bildpunkt des Ausgabebildes ist eine Funktion der bildpunkte in einer wohldefinierten lokalen Umgebung U um den entsprechenden Punkt. (z.B. Filtermatrizen)

Globale Operation: Ein Bildpunkt des Ausgabebildes ist eine Funktion aller Punkte des Eingabebildes bzw. der Eingabebildfolge (z.B. Fouriertransformation, …)

Punktoperatoren

Maximierung des globalen Kontrasts

w ist der Wertebereich ($w_{max} = 255$)

$$g'(g) = \frac{(g - g_{min})(w_{max} - w_{min})}{g_{max} - g_{min}}$$

Quantisierungseffekte durch eine geringe Anzahl an Grauwerten beeinflusst Bild weniger als ein zu geringer Helligkeitsunterschied zwischen ähnlichen Grauwerten. Grauwertverteilung findet keine Berücksichtigung. Grenzen durch einen festen Prozentsatz festlegen (z.B. 0,5%).

Lokaler Kontrast: durchschnittlicher grau Wert zwischen benachbarten Pixeln (keine Punktoperation mehr!)
Verbesserung des lokalen Kontrast durch Log-Transformation: Im Endeffekt eine Gamma-Korrektur

Normiertes Histogramm

Ein normiertes Histogramm beschreibt für jeden Grauwert g die Wahrscheinlichkeit, dass ein beliebiges Pixel diesen Grauwert hat. Entropie ist maximal falls H(f) konstant ist.

Histogrammlinearisierung: Keine Linearisierung, sondern von der Häufigkeit abhängige Spreizung

Adaptive Histogram Equalisation (AHE): Histogramm wird für jeden Punkt für eine vorgegebene Umgebung erstellt. Linearisierung nach diesem Histogramm

Contrast Limited Adaptive Histogram Equalisation (CLAHE): Wie AHE, aber Kontrastverstärkung nur bis zu einem gewissen Maximum

Weitere Punktoperatoren

Farbtransformation: 3 Farbauszüge einer Szene, Farbtransformation durch Wahl geeigneter skalarer Gewichte

$$f_A(i,j) = a \cdot f_R(i,j) + b \cdot f_G(i,j) + c \cdot f_B(i,j)$$

Hintergrundsubtraktion: Aufnahme eines Hintergrundbild H, das die relevanten Objekte des zu verarbeitenden Bildes E nicht enthält
Maskierung: Extraktion semantisch bedeutsamer Teile mit Binärmaske, in der interessante Punkte durch $f(i,j) = 1$ gekennzeichnet sind
Geometrische Transformationen: Transformation der Ortskoordinaten zur Größenanpassung
Alpha-Blending: Methode, um zwei Bilder transparent zu überblenden, α bestimmt Transparenz

Zusammenfassung Punktoperatoren:

- Punktoperatoren finden breite Anwendung zur globalen Bildverbesserung
- Histogramme bieten gute Basis zur Kontrasterhöhung
- Grauwertransformationen (Punktoperatoren) sind jedoch nicht in der Lage, die räumlichen Beziehungen der Grauwerte einer kleinen Umgebung zu erkennen.

Kapitel 3 - Bildverbesserung und Bildanalyse

Flächenbasierte Bildverbesserung (Lokale Operationen)

- Rauschen kann durch Integration einer Signalfolge mit (nahezu) konstantem Signal reduziert werden.
- Konstante Signalfolge: Integration über eine zeitliche Folge oder homogene Fläche
- Lineare verschiebungsinvariante Operatoren (Faltungsmethoden, Filterung im Frequenzraum)
- Beachtung des Randbereichs: konstanter Wert, spiegeln, kopieren

Zeitliche Folge:

- Annahme: Rauschen über mehrerer Bilder über einen gegebenen Zeitraum (normalverteilt)
  Lösung: kann durch Mittelwertbildung

Mittelwertbildung über die Fläche

- Annahme: benachbarte Bildpunkte haben den gleichen Grauwert
  Lösung: Mittelwertbildung über vorgegebene Nachbarschaft (3x3 Boxcar-Filter: 1/9 in jedem Feld)

Min- und Max-Filter

- ersetzen die Werte in einer (z.B. 3x3) Umgebung mit dem enthaltenen minimalen bzw. maximalen Wert (Min-Filter eleminiert weiße Punkte)
- Medianfilter entsprechend

Filteroperationen

- Glättungsoperatoren (Mittelwertfilter, Gauß)
- Kantendetektoren (Prewitt, Sobel, Roberts, Laplace)
- Kombinierte Filter: Laplacian of Gauß

0	1	2	1	0
1	3	5	3	1
2	5	9	5	2
1	3	5	3	1
0	1	2	1	0

Gauß-Glättungsfilter

0	0	-1	0	0
0	-1	-2	-1	0
-1	-2	16	-2	-1
0	-1	-2	-1	0
0	0	-1	0	0

Laplace-Differenzfilter

$$F_{Gauß} = \frac{1}{16}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

Filtermatrizen

Prewitt I $p_x = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$ Prewitt II $p_y = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ Prewitt III $M \approx \sqrt{p_x^2 + p_y^2}$

Sobel I $s_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$ Sobel II $s_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}$ Sobel III $M \approx \sqrt{s_x^2 + s_y^2}$

Roberts-Filter $R_x = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, R_y = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ $R(g(x,y)) = |R_x(g(x,y))| + |R_y(g(x,y))|$

Kompassfilter

Prewitt-Operator (Rotation um 45°)

$$h_1 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix}, \quad h_2 = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix}, \quad h_3 = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \ldots \quad h_8 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -1 \end{pmatrix}$$

Kirschoperator (Rotation um 45°)

$$h_1 = \begin{pmatrix} 3 & 3 & 3 \\ 3 & 0 & 3 \\ -5 & -5 & -5 \end{pmatrix}, \quad h_2 = \begin{pmatrix} 3 & 3 & 3 \\ -5 & 0 & 3 \\ -5 & -5 & 3 \end{pmatrix}, \quad h_3 = \begin{pmatrix} -5 & 3 & 3 \\ -5 & 0 & 3 \\ -5 & 3 & 3 \end{pmatrix} \ldots \quad h_8 = \begin{pmatrix} 3 & 3 & 3 \\ 3 & 0 & -5 \\ 3 & -5 & -5 \end{pmatrix}$$

Laplace I $\nabla^2 \approx \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ Laplace II $\nabla^2 \approx \begin{pmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{pmatrix}$

Laplace ist gegen Rauschen sehr empfindlich, daher davor Glättung mit Gauß-Filter durchführen

LoG II (Mischung aus Laplace und Gauss) $\nabla^2 G(x,y) = \begin{pmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 0 \\ -1 & -2 & 16 & -2 & -1 \\ 0 & -1 & -2 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{pmatrix}$

Kapitel 4 - Bildanalyse und Videoschnitt

Analyse unbewegter Bilder: Farbe, Textur, Kantenbildähnlichkeit, Kanten, Objekte

Segmentierung: Zerlegung eines Bildes in semantische Einheiten.
Segmente: Träger der Bedeutung von Strukturen eines Bildes.

Eigenschaften einer Segmentierung

- *vollständig*: Jedes Pixel ist einem Segment zugeordnet
- *überdeckungsfrei*: ein Pixel ist genau einem Segment zugeordnet
- *zusammenhängend*: jedes Segment bildet ein zusammenhängendes Gebiet.

Ähnlichkeitsbasierte Bildsuche:

- Probleme bei *manuellen Indikatoren*: Vollständigkeit, Kosten im Hinblick auf Zeitaufwand
- *Automatischer Vergleich*: verwendet Merkmale zur Charakterisierung eines Bildes (Vektoren)

Bewegungsanalyse

- Moderne Kompressionsalgorithmen für Video berechnen Bewegungsvektoren für Pixelblöcke (z.B. MPEG-1, ...)
- Bewegungsvektoren sind ohne teure Kalkulationen erhältlich, wenn Encoder/Decoder-Information verwendet wird
- **Optischer Fluss**: Bewegung von grauwertigen Mustern über Bildfläche -> kein verlässliches Merkmal für Inhaltsanalyse!

Schnitterkennung

Trenne Video in Aufnahmen und berechne Merkmale für Aufnahmen getrennt (Algorithmen: Edge Change Ratio, Farbhistogramm, kantenorientierter Kontrast)

- **Histogramm basierte Erkennung**: Schnitt, wenn Farbhistogramme benachbarter Bilder (i-1) und i sich mehr unterscheiden als Schwelle T
  Probleme: verschiedene Bilder können ähnliche Histogramme haben, Explosionen, wechsel

der Beleuchtung werden fälschlicherweise als Schnitt erkannt
Fazit: 90%-98% Performanz, Verbesserungsmöglichkeit durch Einteilung in Regionen

- **Edge Change Ratio**: $ECR_{i-1} = \max\left(\frac{E_{in}}{s_{i-1}}, \frac{E_{out}}{s_i}\right)$
  E_{out}: Pixel in Bild (i-1) ist Kantenpixel, Pixel in Bild i ist kein Kantenpixel
  E_{in}: Pixel in Bild (i-1) ist kein Kantenpixel, Pixel in Bild i ist Kantenpixel
  Problem: Schnelle Objekt- und Kamerabewegung führt zu höheren ECR-Werten ohne Schnitte, Ein-/Aus-/Überblendung schwer zu erkennen
  Fazit: 90%-95% Performanz
- **Ergebnis**:
  - für Schnitte Histogramm
  - für Überblendung Standardabweichung

Treffer-Performanz

	Histo-gramme	ECR	Standard-abweichung	Kantenbas. Kontrast
Schnitt	90-98%	69-97%	n/a	n/a
Überblendung	n/a	n/a	97-100%	n/a
Ausblendung	n/a	0-71%	n/a	16.7-100%

Fehler

	Histo-gramme	ECR	Standard-abweichung	Kantenbas. Kontrast
Schnitt	5-20%	9-17%	n/a	n/a
Überblendung	n/a	27-700%	8-57%	n/a
Ausblendung	n/a	47-37000%	n/a	20-3000%

Erkennung des Kamerabetriebs

- Verwende Bewegungsvektoren zur Kompression von Algorithmen oder berechne optischen Fluss, um Bewegungsvektoren eines Videos zu erhalten
- Teste, ob Bewegungsvektoren vordefinierten Kamerabetriebsmustern im Hinblick auf absolute Länge und Orientierung entsprechen
  - die meisten Vektoren sind mit derselben Ausrichtung -> Schwenk
  - konzentrische Vektoren -> Zoom
- Problem: funktioniert bei Objektbewegung kaum

Kapitel 5 - Video- und Audioanalyse

Szenenanalyse

Schlüsselbilder (Key-Frames)

- stellen den Inhalt einer Szene in komprimierter Form dar
- benötigen semantisches Verständnis (z.B. Gesichtserkennung) einer Szene
- Auswahl geeigneter Eigenschaften zur Detektion: Farbe, Textur, Bewegung von Bild zu Bild

Erkennen eines Videogenres

- Verbindung syntaktischer und semantischer Merkmale mit charakteristischem Fingerabdruck
- Erstelle Datenbasis mit Fingerabdrücken
- Springe zum nächsten harten Schnitt und berechnende den Fingerabdruck der nächsten Sekunden und vergleiche mit der Datenbasis

Videoabstraktion

- **Ziel**: Automatische Erzeugung eines Videotrailers aus einem ganzen Video
- **Vorgehensweise**:
  - Trennung von Videoteilen durch Schnitterkennung
  - Charakterisierung der Aufnahmen durch existierende Werkzeuge
  - Definition der Heuristik, um nötige Teile mit einzubeziehen im Hinblick auf die Reihenfolge im Trailer

Syntaktische Audiomerkmale

- **Lautstärke**: kann direkt von darunter rohen Audiodaten erhalten werden
- **Frequenzverteilung**: Ergebnis einer Fourier-Transformation von rohen Audiodaten
- Psychoakustische Eigenschaften sind komplex
- Gefühl der Laustärke nicht wirklich messbar

Audioschnitterkennung

- Problem: äußerst schwierig (Geräusche, Musik, Sprache, ...)
- Lösung: verwende Merkmalsvektor für Zeitfenster i (Lautstärke, Frequenz, Tonhöhe) und vergleiche diese Zeitfenster, falls Differenz vorhanden dann war es ein Schnitt (Schwellwert)

Definitionen

- **Onset**: Zunahme von Feuerwahrscheinlichkeit
- **Offset**: Abnahme von Feuerwahrscheinlichkeit
- **Spektrogramm**: Darstellung des zeitlichen Verlaufes des Frequenzspektrums eines Signals
- **Formant**: Konzentration akustischer Energie in einem bestimmten Frequenzbereich, entstehen z.B. in den Resonanzspektren der Musikinstrumente oder der menschlichen Sinne

Audioanalyse

Erkennen von Stille

- Erkennen von Intervallen relativer Stille in einem Audio-Track
- Vorgehensweisen
    - Maße der Lautstärke: Partien werden als relative Stille wahrgenommen
      Problem: Festlegen der Lautstärkenschwelle
    - Energiemaße werden in Kombination mit dem Nulldurchgang angewendet, um Wortgrenzen zu identifizieren

Zeitliche Struktur

- **Takt**: Amplitudenstatistik liefert den Takt
- **Rhythmus**: Frequenzdomäne liefert den Rhythmus

Anfrage nach ähnlichen Musikstücken

- Ähnliche Musikstücke können durch Fingerprints (Shazam) realisiert werden.
- Aus Spektrogramm lokale Maxima (Peaks) erkennen und anhand dieser vergleichen
- Abweichende können auch Störgeräusche sein
- Map aus Peaks wird partiell gehasht um schnell vergleichen zu können

Kapitel 6 - Lokalisierung und Navigation

Globale Navigationssysteme

- Ein Beispiel hierzu ist Galileo

Lokalisierungsverfahren - Signalfeldstärke

$$\alpha = \varphi + \varepsilon$$

$$\varepsilon = (t_1 - t_2) \cdot \omega$$

$$\Rightarrow \alpha = \varphi - (t_2 - t_1) \cdot \omega$$

Mathematisches Lösen einer Navigationsgleichung

- Annahme: N Landmarken an verschiedenen bekannten Positionen (x_i, y_i, z_i) mit den Entfernungsnagaben r_i zu einer zu bestimmenden Position (x,y,z)
- $$(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2 = r_i^2$$

Lokalisierung in Gebäuden

System	Medium	Verfahren	Genauigkeit
Active Badge	Infrarot	Cell-ID	Raumgröße / 100%
Active Office	Radiowellen, Ultraschall	TDoA	8 cm / 95%
Cricket	Radiowellen, Ultraschall	TDoA	1,2 m / 100%
RADAR	WLAN	RSSI	3 m / 95%
Overlapping Connectivity	Radiowellen	-	1/3 *(Ankerabstand) / 90%
APIT	Radiowellen	RSSI	Größe der Überschneidungen / 85%
Ultra-Wideband	Richtfunk	TDoA/AoA	15 cm / 50%
BT-Localize	Bluetooth	RSSI	1,7 m / 50% (MMSE) 0,9 m / 50% (Partikelfilter)

Kapitel 7 - Lokalisierung

Vorstellung Sensoren in mobilen Endgeräten

- Multi-Touch Display
- GNSS-Empfänger (Lokalisierung)
- Beschleunigungssensor (Bewegung)
- Magnetischer Kompass (Orientierung)
- RFID/NFC-Lesegerät (Transponder), ...

Mustererkennung mit Dynamic Time Warping

- Annahme dass alle Signale aus diskreten Datenpunkten bestehen
- Sucht nach dem optimalen Pfad mit minimalen Kosten
- Kosten sind die Summe der Abstände zwischen den passenden Datenpunkten.

Sensordatenfusion – Koppelnavigation

- **Ziel**: Schätzung der aktuellen Position aufgrund aktueller Sensormessungen und vorheriger Position

Zum Beispiel: Partikel als Positionsschätzer

Bayes-Filter V - Implementierung als Partikelfilter

- Idee: Repräsentiere Bel(x) durch eine Menge von m gewichteten Samples
- $Bel(x) = \{x^{(i)}, p^{(i)}\}_{i=1,\dots,m}$ mit $p^{(i)} \geq 0;\ \sum p^{(i)} = 1$
- x Sample und p Gewicht

Kapitel 8 - Gesten- und Aktivitätserkennung durch Tiefenbildinformationen

Vorteile einer Time-of-Flight-Kamera

- Nur eine (spezielle) Kamera erforderlich
- Keine manuelle Tiefenberechnung notwendig
- Erfassen von 3D-Szenen in Echtzeit
- Abgeschwächte Abhängigkeit von der Szenenbelechtung

Pulsmodulation

- **Vorteile**:
  - Direkte Messung der Signallaufzeit
  - Hochenergetische Lichtimpuls reduziert Abhängigkeit von Hintergrundbeleuchtung
  - Beleuchtungs-und Beobachtungsrichtung sind zueinander kollinear
- **Nachteil**:
  - Hochpräzise Zeitmessungen erforderlich
  - Messung der reflektierten Lichtimpulse ist aufgrund von Streuungen inexakt
  - Schwierigkeit kurze Lichtimpulse mit steilen Signalflanken zu generieren

Continuous Wave Modulation

- Ermittlung der Phasenverschiebung durch Demodulation
- Demodulation durch Kreuzkorrelation des emittiertem mit dem empfangenen Signal
- **Vorteile**:
  - Vielzahl geeigneter Lichtquellen
  - Auf verschiedene Modulationsverfahren anwendbar
  - Gleichzeitiges Entfernungs- und Amplitudenbild
- **Nachteile**:
  - In der Praxis ist eine Integration über die Zeit erforderlich, um Rauschen zu reduzieren
  - Bildraten begrenzt durch Integrationszeit
  - Bewegungsunschärfe durch lange Integrationszeiten

Funktionsprinzip der Kinect

Streifenprojektion: Man projiziert ein definiertes Lichtmuster in eine reale 3D-Szene, Verzeichnung der eingeprägten Lichtmuster erlaubt Rekonstruktion der 3D-Struktur

Arten der Streifenprojektion: Einzigeartige Kodierung für jede Bildposition, damit jeder Punkt im Muster eindeutig zugeordnet werden kann. (Pseudozufallsmuster)

Kapitel 9 - Dynamic Time Warping - ein Algorithmus, viele Anwendungen

Vertiefung: Dynamic Time Warping

- Mustervergleich zweier Zeitsequenzen mit variierenden Geschwindigkeit
- Berechnung mittels dynamischen Programmierens (zerlegen in Teilprobleme und Wiederverwendung zur Berechnung der nächstgrößeren Teillösung)
- **Anwendungen**: Editierdistanzen, Synchronisierung von Musik, Spracherkennung, Gestenerkennung, Data Mining, Bioinformatik

Levenshtein-Distanz

- Damerau-Levenstein-Distanz: berücksichtigt die Vertauschung von 2 Buchstaben
- Modifizierte Kostenfunktion:
  Berücksichtigung der Tastaturanordnung
  Kosten für phonetisch gleich klingende Buchstaben geringer gewichten

Editierdistanz $d(i,j)$ für Teilwörter $A = a_1, \dots, a_i$ und $B = b_1, \dots, b_j$

$d(0,j) = j$ (komplettes Wort erzeugen)
$d(i,0) = i$ (komplettes Wort löschen)

$$d(i,j) = \min(\, d(i,j-1)+1, \text{ (Löschen)}$$
$$d(i-1,j)+1, \text{ (Einfügen)}$$
$$d(i-1,j-1)+1, \text{ falls } a_i \neq b_j \text{ (Ersetzen)}$$
$$d(i-1,j-1)+0, \text{ falls } a_i = b_j \text{ (Übernehmen)}$$
$$)$$