

Inhaltsbasierte Bildersuche

Medieninformatik IL

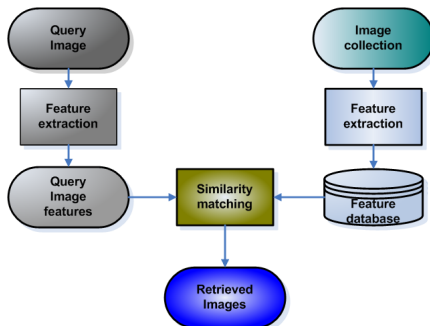
Andreas Unterweger

Vertiefung Medieninformatik
Studiengang ITS
FH Salzburg

Wintersemester 2018/19

Definition „Inhaltsbasierte Bildersuche“

- Engl. *Content-based image retrieval (CBIR)*
- Finden von Bildern anhand von Bildmerkmalen
 - Farben und deren Verteilung
 - Formen und deren Verteilung
 - Merkmalsvektoren (z.B. von SIFT)
 - ...



Quelle: Rorissa, A.: Image Retrieval – Benchmarking Visual Information Indexing and Retrieval Systems.
<http://www.asis.org/Bulletin/Feb-07/rorissa.html> (4.10.2014), 2007.

- Motivation: Weitestgehende Vermeidung manueller Metadatenpflege (z.B. durch Tagging), da
 - Zeitaufwändig
 - Kaum automatisierbar
 - Umfangsbeschränkt (z.B. konsistente Farbsortierung kaum möglich)
- Hauptanwendung: Ähnlichkeitssuche (engl. *Reverse Image Search* bzw. *Query by Example*)
 - Finden von Bildern mit ähnlichen Farben/Motiven/...
 - Duplikaterkennung
 - Plagiaterkennung
 - Manipulationserkennung
- Andere Anwendungen (Auswahl):
 - Semantische Suche (Finden bestimmter Motive)
 - Zensur (Hauterkennung u.ä.)

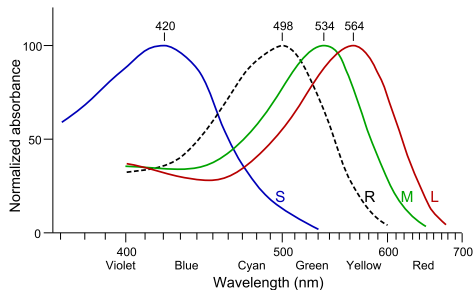
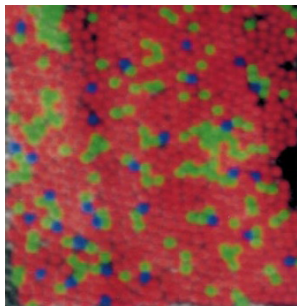
- Google Bildersuche (<https://www.google.at/imghp>)
- TinEye (<http://tineye.com/>)
- pixolution (<http://demo.pixolution.de/>)
- The GNU Image-Finding Tool
(<http://www.gnu.org/software/gift/>)
- Übersichtsliste:
http://en.wikipedia.org/wiki/List_of_CBIR_engines

- Merkmaltaxonomie nach¹
 - **Low-level: Aus Pixeldaten**
 - Mid-level: Aus segmentierten Daten
 - High-level: Semantische Daten (Kategorien, Zusammenhänge etc.)
- *Semantic Gap*: Mensch sucht high-level, Maschine arbeitet low-level
- Ausgewählte Bildmerkmale
 - Durchschnittsfarbe(n)
 - SIFT-Merkmalvektoren
 - Farbverteilung
- Ausgewählte Vergleichsmetriken
 - Vektordistanz
 - Histogrammdistanz

¹Marques, O. und Furht, B.: Content-Based Image and Video Retrieval. Kluwer Academic Publishers, 2002.

Wiederholung: Menschliche Farbwahrnehmung

- Farbreiz aus Rot-, Grün- und Blau-Komponenten (RGB)
- Drei Arten von Zapfen auf Netzhaut → Farbwahrnehmung
- Andere Farbräume zur Berücksichtigung der weiteren Verarbeitung



Quellen: Roorda, A. und Williams, D.: The arrangement of the three cone classes in the living human eye. *Nature*, 397(6719):520-522, 1999; <http://commons.wikimedia.org/wiki/File:Cone-response.svg>

Durchschnittsfarbe(n) I

- Ausgangspunkt: $m \cdot n$ Pixel großes RGB-Bild $I(x, y) = \begin{pmatrix} I_R(x, y) \\ I_G(x, y) \\ I_B(x, y) \end{pmatrix}$
- Berechnung: Mittelwert jedes Farbkanals:

$$\bar{I}_R = \frac{1}{m \cdot n} \cdot \sum_{y=1}^m \sum_{x=1}^n I_R(x, y)$$

$$\bar{I}_G = \frac{1}{m \cdot n} \cdot \sum_{y=1}^m \sum_{x=1}^n I_G(x, y)$$

$$\bar{I}_B = \frac{1}{m \cdot n} \cdot \sum_{y=1}^m \sum_{x=1}^n I_B(x, y)$$

$$\rightarrow \bar{I} = \begin{pmatrix} \bar{I}_R \\ \bar{I}_G \\ \bar{I}_B \end{pmatrix}$$

$$\bar{I} = \begin{pmatrix} \bar{I}_R \\ \bar{I}_G \\ \bar{I}_B \end{pmatrix}$$

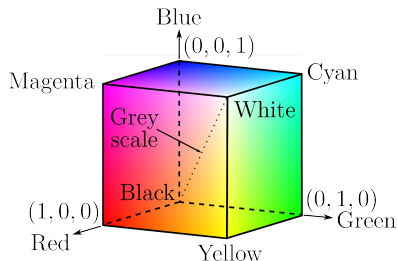
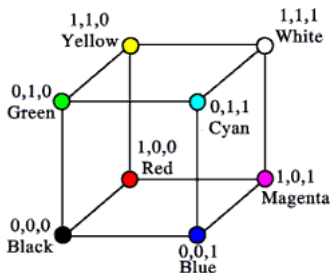
$$\bar{I} = \frac{1}{m \cdot n} \cdot \sum_{y=1}^m \sum_{x=1}^n I(x, y)$$

- Vorteile:
 - Einfach zu berechnen
 - Schnell zu berechnen
- Nachteile:
 - **Helligkeitsabhängig**
 - Kaum diskriminativ (unähnliche Bilder liefern ähnlichen Durchschnitt)
 - Nicht repräsentativ (rot-blaues Muster gleicht lila Fläche)

Durchschnittsfarbe(n) III

- Helligkeitsabhängigkeit durch Normalisierung entfernen
- Richtung (Farbkomponentenaufteilung) bleibt erhalten, Summe wird 1

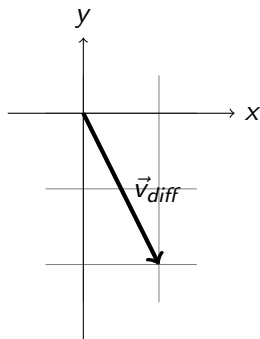
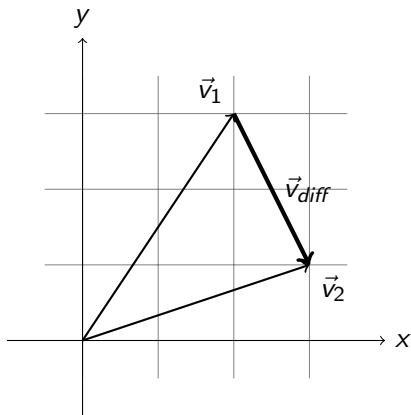
$$\bar{I}_n = \frac{\bar{I}}{\bar{I}_R + \bar{I}_G + \bar{I}_B} \left(= \frac{\bar{I}}{|\bar{I}_R| + |\bar{I}_G| + |\bar{I}_B|} = \frac{\bar{I}}{\|\bar{I}\|_1} \right)$$



Quellen: Otto, G.: Organizing Color. http://viz.aset.psu.edu/gho/sem_notes/color_2d/html/primary_systems.html (4.10.2014), 2000; Cattin, P.: Digital Image Fundamentals – Introduction to Signal and Image Processing. <https://miac.unibas.ch/SIP/02-Fundamentals.html> (4.10.2014), 2014.

Vektorabstand I

- Abstand zwischen zwei Vektoren \vec{v}_1 und \vec{v}_2 bestimmen
- Typisches Maß: Norm des Differenzvektors $\vec{v}_{diff} = \vec{v}_2 - \vec{v}_1$



- p -Norm im \mathbb{R}^n für einen Vektor $\vec{v} = (v_1 \ v_2 \ \dots \ v_n)^T$:

$$\|\vec{v}\|_p = \left(\sum_{i=1}^n |v_i|^p \right)^{\frac{1}{p}}$$

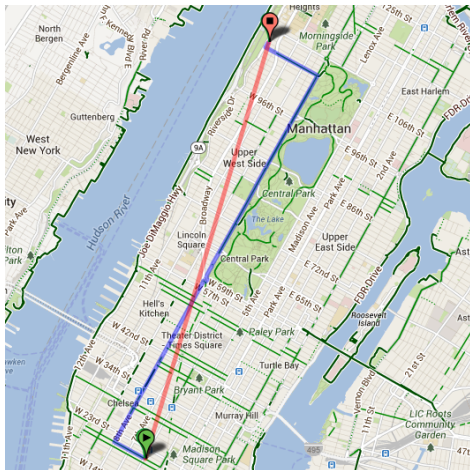
- Praktisch relevanter Spezialfall 2-Norm (euklidische Norm):

$$\|\vec{v}\|_2 = \sqrt{\sum_{i=1}^n v_i^2}$$

- Praktisch relevanter Spezialfall 1-Norm (Summennorm):

$$\|\vec{v}\|_1 = \sum_{i=1}^n |v_i|$$

- Summennorm wird auch als Manhattan-Distanz bezeichnet:



Quelle: Harmony Institute: Comparing the moral language of television to that of its viewers. <http://harmony-institute.org/therippleeffect/2013/06/26/comparing-the-moral-language-of-television-to-that-of-its-viewers/> (4.10.2014), 2013.

- Vektorabstand auch für Merkmalvektoren ($\vec{v}_1, \vec{v}_2 \in \mathbb{R}^n$) berechenbar
- Beispiel SIFT: Schlüsselpunktdeskriptoren sind 128-dimensionale Merkmalvektoren ($\vec{v}_1, \vec{v}_2 \in \mathbb{R}^{128}$)

→ Norm des Differenzvektors ergibt Abstand

- Abstände von nicht normierten Vektoren sind oft problematisch
 - Stärkere Gewichtung des längeren Vektors
 - Kurze Vektoren in entgegengesetzte Richtungen haben kurze Abstände

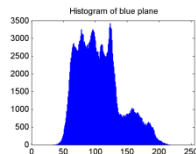
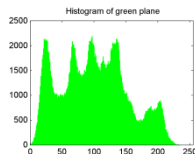
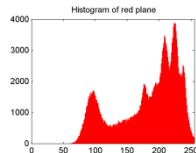
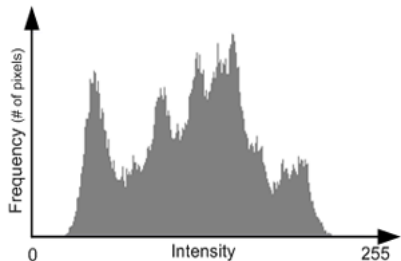
→ Winkel zwischen Vektoren als alternatives Ähnlichkeitsmaß

$$\varphi = \arccos \left(\frac{\vec{v}_1 \cdot \vec{v}_2}{\|\vec{v}_1\| \cdot \|\vec{v}_2\|} \right)$$

- Anmerkung: \cdot bezeichnet im Zähler das (innere) Vektorprodukt

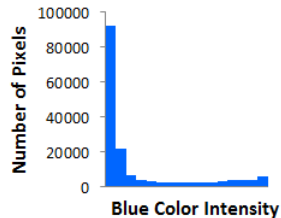
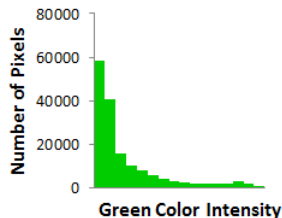
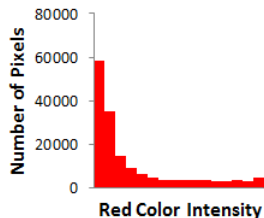
Farbverteilung I

- Farbverteilung über Histogramme jedes Farbkanals ermittelbar
- Histogramm: Häufigkeit von Farbwerten oder Farbwertgruppen



Adaptiert von: Gemma, R. M.: Image histogram. <http://technlg.wordpress.com/2014/01/23/image-histogram/> (4.10.2014), 2014.

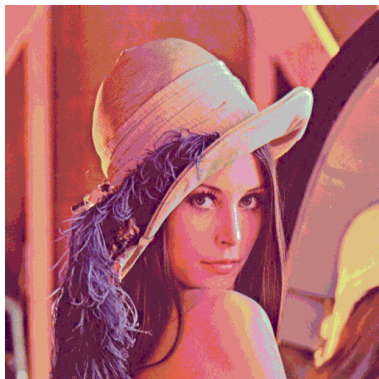
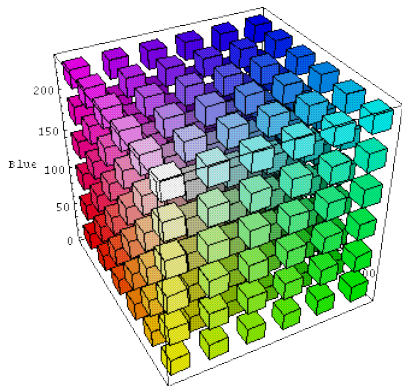
- Gruppierung von Farbwerten durch (Vektor-)Quantisierung



Quelle: Cloud'N'Sci Ltd: RGB Heatmap Demo. <http://cloudnsci.fi/wiki/index.php?n=HeatMiner.RgbHeatmapDemo> (4.10.2014), 2013.

- Normierung sinnvoll
- Verwendung anderer Farbräume möglich

- Quantisierungseffekte:



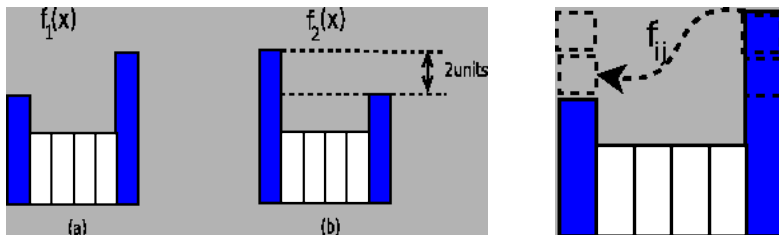
Quelle: Ji, J.: Chapter 2. Image Structure <http://jun.hansung.ac.kr/DI/Chapter-2.htm> (4.10.2014), 2014.

- Vergleich von Histogrammen durch Histogrammdistanz (entweder pro Farbkanal separat oder durch zusammengefügte Histogramme)
- Verschiedene Metriken zur Distanzbestimmung (Auswahl):
 - **Minkowski-Distanz** (p -Norm)
 - Kreuzkorrelation
 - **Earth Mover's Distance (EMD)**
- Minowski-Distanz d_p für Histogramme H_1 und H_2 mit je N Partitionen (Partitionsmenge \mathbb{P} , $|\mathbb{P}| = N$, $H_i : \mathbb{P} \rightarrow [0; 1] \cap \mathbb{R}$):

$$\|H_{diff}\|_p = \left(\sum_{p' \in \mathbb{P}} |H_2(p') - H_1(p')|^p \right)^{\frac{1}{p}} = \left(\sum_{i=1}^N |H_{2_i} - H_{1_i}|^p \right)^{\frac{1}{p}}$$

Earth Mover's Distance I

- Idee: Histogramme repräsentieren Gelände (Häufigkeit jeder Partition entspricht Höhe eines Erdhügels $\rightarrow N$ Erdhügel)
- Frage: Wie viel Arbeit ist (minimal) notwendig, um mit einer Schürfrau das erste Gelände in das zweite umzuformen?
- Ziel: Aufwand für Umformarbeiten entspricht Histogrammdistanz
- Voraussetzung: Histogramme sind normiert („gleich viel Erde“)



Quelle: Čonos, M.: Earth Mover's Distance – EMD.

<http://vellum.cz/~mikc/oss-projects/CarRecognition/doc/dp/node29.html> (4.10.2014) 2007.

Earth Mover's Distance II

- Definition Arbeit: Verschobene Menge f mal Entfernung d zwischen Partitionen mit Indizes i und j :

$$W_{ij} = f_{ij} \cdot d_{ij}$$

- Arbeit für alle Verschiebungen (soll durch f_{ij} -Wahl minimiert werden):

$$W = \sum_{i=1}^N \sum_{j=1}^N W_{ij} = \sum_{i=1}^N \sum_{j=1}^N f_{ij} \cdot d_{ij}$$

- EMD als Metrik: Normierung der Arbeit durch verschobene Menge F :

$$EMD(H_1, H_2) = \frac{W}{F} = \frac{\sum_{i=1}^N \sum_{j=1}^N f_{ij} \cdot d_{ij}}{\sum_{i=1}^N \sum_{j=1}^N f_{ij}}$$

Fragen?