VK Multimedia
Information Systems

Mathias Lux, mlux@itec.uni-klu.ac.at

Dienstags, 16.00 Uhr c.t., E.2.69

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 2.0 License. See http://creativecommons.org/licenses/by-nc-sa/2.0/at/
Projects for consideration

- Video Summarization
  - Finding the most important frames of a short video
- Java Face Detection
  - Finding faces in arbitrary images
- Cross-Platform QT 4.4 Media Player
  - Experiences ...
- Automated Image Quality Measurements
  - Blurriness, block artifacts, etc.
- Implementation of Image Features
  - E.g. region based (Blobworld)
- Non-Implementational
  - Subject to discussion ...
Content Based Image Retrieval

- Motivation & Semantic Gap
- Perception
- Color Based Features
- Texture Based Features
- Segmentation
- Object Recognition
- Vector Images
- Evaluation
Motivation

Lots of good reasons ...

● Visual information overload
  • Devices (cameras, mobile phones, etc.)
  • Communication (email, mo-blogs, etc.)

● Metadata not available
  • Time consuming
  • No automation
Question: What is so special ‘bout Mona Lisa’s smile?
Semantic Gap

● Also called **Sensory Gap**

● Defined as
  
  • Inability of automatic understanding
  
  • Gap between high- and low-level features / metadata

● Actually hard task for humans also
Semantic Gap (1)

- General Definition: Santini & Jain (1998)
Where actually is the Semantic Gap?

- Classification based on Concepts
- Segmentation & Object Recognition
- ...
Applications

● **Home User & Entertainment**
  • Find picture of / from / at ....
  • Search & browse personal digital library

● **Graphics & Design**
  • Find picture representing something (Color in CD/CI, feeling, etc.)

● **Medical Applications**
  • Diagnosis, segmentation & classification
  • X-Ray images, patient monitoring
Applications

● Accessibility
  • ‘Explain’ image to visually disabled people

● Industrial application
  • Select / Sort out products (chips, buns)
  • Monitor processes (e.g. sensors unavailable)

● Security
  • Match fingerprints
  • Search face database
Applications

● Biology
  • Analysing cell samples
  • Recognizing animals, insects & plants

● Astronomy
  • Classifying stars & events

● Weather forecasting
  • Satellite images, clouds

● Cartography
  • Mapping (e.g. aerial photo - earth model)
Content Based Image Retrieval

- Motivation & Semantic Gap
- Perception
- Color Based Features
- Texture Based Features
- Segmentation
- Object Recognition
- Vector Images
- Evaluation
Perception

- The **eye** as instrument of perception
- Sensory capabilities
  - Cones (bright light): 6-7 Mio.
  - Brain ‘corrects’ vision
    - e.g. blind spot
Color & Color Spaces

S-, M- and L-cones: Blue, green and red

- RGB based on these three colors
- CIE models perception better
  - Responsiveness of cone types
  - Number of cones / types
  - etc.
The human eye ...

- Count the black dots on the image:
The human eye ...

- Rabbit or duck?
The human eye ... 

- Anamorphic illusions
What are (digital) images?

● An Image is
  • Created by a set of photons
    • With different frequency
    • Moving from different sources
    • Along different vectors
  • A representation of sensor unit activation
    • Activated by the set of photons

● Storing an image
  • Based on the set of photons
Sampling & Quantization

- Capturing continuous images on sensors
  - Sampling: Continuous to matrix
  - Quantization: Continuous color to value
Sampling & Quantization

- Size of a captured image:
  - \( \text{# of samples (width*height)} \times \text{# of colors} \)
Content Based Image Retrieval

- Motivation & Semantic Gap
- Perception
- Color Based Features
- Texture Based Features
- Segmentation
- Object Recognition
- Vector Images
- Evaluation
Color Histogram

- Count how often which color is used
- Algorithm:
  - Allocate int array \( h \) with \( \text{dim} = \# \text{ of colors} \)
  - Visit next pixel -> it has color with index \( i \)
  - Increment \( h[i] \)
  - IF pixels left THEN goto line 2
- Example: 4 colors, 10*10 pixels
  - histogram: \([4, 12, 20, 64]\)
Color Histogram

● Strategies:
  • Quantize if too many colors
  • Normalize histogram (different image sizes)
  • Weight colors according to use case
  • Use (part of) color space according to domain

● Distance / Similarity
  • Assumption: All images have the same colors
  • $L_1$ or $L_2$ is quite common
Color Histogram

● Benefits
  • Easy to compute, not depending on pixel order
  • Matches human perception quite good
  • Quantization allows to scale size of histogram
  • Invariant rotation & reflection

● Disadvantages
  • Distribution of colors not taken into account
  • Colors might not represent semantics
  • Find quantization fitting to domain / perception
  • Image scaling might be a problem
Color Histogram

- Example: 4 images, 7 colors
  - 1: [0, 4, 12, 20, 64, 0, 0]
  - 2: [66, 4, 12, 20, 0, 0, 0]
  - 3: [0, 0, 0, 64, 0, 20, 16]
  - 4: [0, 0, 0, 0, 64, 20, 16]
Color Histogram

Bilder:
- 1: [0, 4, 12, 20, 64, 0, 0]
- 2: [66, 4, 12, 20, 0, 0, 0]
- 3: [0, 0, 0, 64, 0, 20, 16]
- 4: [0, 0, 0, 0, 64, 20, 16]

Distanzfunktion $d$: L1
- $d(1, 2) = 130$
- $d(1, 3) = 160$
- $d(1, 4) = 52$
Dominant Color

- Reduce histogram to dominant colors
  - e.g. for 64 colors c0-c63:
    - image 1: c12 -> 23%, c33 -> 6%, c2 -> 2%
    - image 2: c11 -> 43%, c2 -> 12%, c54 -> 10%

- Distance function in 2 aspects:
  - Difference in amount (percentage)
  - Difference between colors (c11 vs. c12)

- Further aspects:
  - Diversity and distribution
Dominant Color

● Benefits:
  • Small feature vectors
  • Easily understandable & intuitive
  • Similarity of color pairs (light vs. dark red, etc.)
  • Invariant to rotation & reflection

● Disadvantages
  • Similarity of color pairs no trivial problem
  • Colors might not represent semantics
  • Find quantization fitting to domain / perception
Color Distribution

- Index dominant color in image segment
  - e.g. 8*8 = 64 image segments
  - feature vector has 64 dimensions
    - One for each segment
  - color index is the entry on segment dimension
    - e.g. 16 colors [2, 0, 3, 3, 8, 4, ...]
Color Distribution

● Similarity
  • $L_1$ or $L_2$ are commonly used

● Benefits
  • Works fine for many scenarios
    • clouds in the sky, portrait photos, etc.
  • Mostly invariant to scaling

● Disadvantages
  • Colors might not represent semantics
  • Find quantization fitting to domain / perception
  • Rotation & reflection are a problem
Color Correlogram

● Histogram on
  • how often **specific colors** occur
  • in the **neighbourhood** of each other
● Histogram size is (\# of colors)^2
  • For each color an array of neighboring colors
Color Correlogram

- Extraction algorithm
  - Allocate array $h[#\text{colors}][#\text{colors}]$ all zero
  - Visit next pixel $p$
  - For each pixel $q$ in neighborhood of $p$:
    - increment $h[\text{color}(p)][\text{color}(q)]$
  - IF pixels left THEN goto line 2

- Algorithm is rather slow
  - Depends on size of neighborhood
  - Typically determined by city block distance
Color Correlogram

● Similarity
  • $L_1$ or $L_2$ are commonly used

● Benefits
  • Integrates color as well as distribution
  • Works fine for many scenarios
  • Mostly invariant to rotation & reflection

● Disadvantages
  • Find appropriate neighborhood size
  • Find quantization fitting to domain / perception
  • Rather slow indexing / extraction
Color Correlogram

- Auto Color Correlogram
  - Just indexing how often $color(p)$ occurs in neighborhood of pixel $p$
  - Simplifies the histogram to size # of colors
Color Correlogram

- Integrating different pixel features to correlate
  - **Gradient Magnitude** (intensity of change in the direction of maximum change)
  - **Rank** (intensity variation within a neighborhood of a pixel)
  - **Texturedness** (number of pixels exceeding a certain level in a neighborhood)
Content Based Image Retrieval

- Motivation & Semantic Gap
- Perception
- Color Based Features
- Texture & Shape Features
- Segmentation
- Object Recognition
- Vector Images
- Evaluation
Texture & Shape Features

- Indexing non color features in image
  - Outlines, edges of regions
  - Overall characteristics like coarseness and regularity
Spatial Filtering

- Methods for *enhancing* the image
- Normally a kernel or filter is used:
  - A matrix which is applied to the image
  - In a linear transformation
## Spatial Filtering

### Table

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>194</td>
<td>128</td>
<td>102</td>
<td>197</td>
<td>69</td>
</tr>
<tr>
<td>162</td>
<td>68</td>
<td>103</td>
<td>144</td>
<td>115</td>
</tr>
<tr>
<td>121</td>
<td>85</td>
<td>57</td>
<td>27</td>
<td>14</td>
</tr>
<tr>
<td>24</td>
<td>183</td>
<td>192</td>
<td>239</td>
<td>150</td>
</tr>
<tr>
<td>92</td>
<td>93</td>
<td>154</td>
<td>138</td>
<td>170</td>
</tr>
</tbody>
</table>

### Diagram

The diagram illustrates the concept of spatial filtering, showing how the values in the table are adjusted by applying a smoothing filter. The green cells indicate the central pixel and its neighbors, which are used to calculate the new value for the central pixel. The values are then updated according to the filter coefficients shown in the smaller table on the right.
Spatial Filtering

- This is a simple smoothing kernel
- Other operations
  - Sharpen
  - Unsharp Mask
  - Gradient
Edge Detection

- Based on the gradient
  - Denotes the amount of change at specific point
  - Can be estimated with Gradient $\nabla f \approx |G_x| + |G_y|$
- Different kernels based on estimation
  - Test e.g. with Gimp
Edge & Texture Features

• Edge based on gradient map
  • Aims to represent edges in number vector
  • e.g. Edge Histogram (MPEG-7)
  • Problems with rotation & reflection

• Texture features
  • Statistics on the image representing
  • Heavily depends on domain
  • Mostly invariant to rotation & reflection
  • Problems with scaling
Tamura Features
Tamura & Mori (1978)

- Widely used in CBIR
  - E.g. IBM QBIC
- 6 texture features
  - Coarseness, contrast, directionality
  - Line-likeness, regularity, and roughness
- Good overview is provided in:
Tamura Features
Tamura & Mori (1978)

- **Coarseness**
  - Pixel diversity in neighborhoods
- **Contrast**
  - Using mean and variance of an image
- **Directionality**
  - Horizontal and vertical derivatives (like Sobel)
Shapes

- Indexing of
  - **Boundaries** and **Regions**
  - Invariant to scaling, rotation and translation
- Features depend on domain, e.g.
  - Length of outline w.r.t. the image
  - Convexity & Concavity
  - Holes & Connectivity
Content Based Image Retrieval

- Motivation & Semantic Gap
- Perception
- Color Based Features
- Texture Based Features
- Segmentation
- Object Recognition
- Vector Images
- Evaluation
Segmentation

- Automatically decompose image
  - Foreground & background
  - Different objects & shapes
- Example applications
  - Computer vision in industry: Counting
  - Security: Erasing background
Application: Blobworld

**Step 1:**

To begin a query, select a blob by clicking in the Blobworld image above.

You can also type in one or more keywords. We’ll search the Corel keywords, caption, and CD title, and only do the Blobworld search among images that match all of your keywords. *(But read this warning about the inaccuracy of keywords.)*

Or search based on keywords alone — just type the keywords and click “Submit.”
Application: Blobworld

Step 2:
Adjust the weights below if you’d like, then click “Submit.”

<table>
<thead>
<tr>
<th>How important is the selected region?</th>
<th>Not</th>
<th>Somewhat</th>
<th>Very</th>
</tr>
</thead>
<tbody>
<tr>
<td>How important are the features of this region?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shape/Size</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How important is the background (everything outside the region)?</th>
<th>Not</th>
<th>Somewhat</th>
<th>Very</th>
</tr>
</thead>
</table>
Application: Blobworld

Querying from 25000 images (2000 returned by the filter).
Application: Blobworld

- Segmentation Method: Pixel clustering
  - Each pixel has feature vector
    - Color, texture in neighbourhood, etc.
  - Disjunctive clusters group similar pixels
- Method available in MatLab source code
Segmentation Example: Region Growing - Flood Fill

- Image is represented by its gradient map
- Some pixel is selected as seed
- Imaginary water is poured onto this seed
  - Flooding pixels with lower or equal height

Problems:
- Dam breach vs. amount of water
- Good seed
Content Based Image Retrieval

- Motivation & Semantic Gap
- Perception
- Color Based Features
- Texture Based Features
- Segmentation
- Object Recognition
- Vector Images
- Evaluation
Object Recognition

- Motivated by practical applications
  - License plate detection & recognition
  - Logo detection & classification (e.g. in TV)
  - Face detection & recognition
  - Identifying spatial regions of interest
    - possible tumors in medical imaging
Object Recognition

Weight of problem depends on domain ... 

- License plate detection is rather easy
  - Plates always look the same
  - Small number of possible positions
  - Sensor fixed, background is separated easily

![Image of a sign with a speed limit and a mobility scooter]

ITEC, Klagenfurt University, Austria – Multimedia Information Systems
## Face Recognition

<table>
<thead>
<tr>
<th>Areas</th>
<th>Specific applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entertainment</td>
<td>Video game, virtual reality, training programs</td>
</tr>
<tr>
<td></td>
<td>Human-robot-interaction, human-computer-interaction</td>
</tr>
<tr>
<td>Smart cards</td>
<td>Drivers’ licenses, entitlement programs</td>
</tr>
<tr>
<td></td>
<td>Immigration, national ID, passports, voter registration</td>
</tr>
<tr>
<td></td>
<td>Welfare fraud</td>
</tr>
<tr>
<td>Information security</td>
<td>TV Parental control, personal device logon, desktop logon</td>
</tr>
<tr>
<td></td>
<td>Application security, database security, file encryption</td>
</tr>
<tr>
<td></td>
<td>Intranet security, internet access, medical records</td>
</tr>
<tr>
<td></td>
<td>Secure trading terminals</td>
</tr>
<tr>
<td>Law enforcement and surveillance</td>
<td>Advanced video surveillance, CCTV control</td>
</tr>
<tr>
<td></td>
<td>Portal control, postevent analysis</td>
</tr>
<tr>
<td></td>
<td>Shoplifting, suspect tracking and investigation</td>
</tr>
</tbody>
</table>
Face Recognition

Input Image/Video

Face Detection

Feature Extraction

Face Recognition

Identification/Verification

Other Applications
- Face Tracking
- Pose Estimation
- Compression
- HCI Systems

Other Applications
- Facial Feature Tracking
- Emotion Recognition
- Gaze Estimation
- HCI Systems

Approaches
- Holistic Templates
- Feature Geometry
- Hybrid
Face Detection

- **Task:**
  - Isolate face from background / other faces

- **Problems:**
  - Partially hidden
  - Point of view

- **Methods:**
  - Templates (whole face & feature based)
  - Skin color
  - Neural networks & machine learning
Feature Extraction

● Extraction of characteristics
  • Different approaches (e.g. Eigenfaces and Fisherfaces)
  • Psychological background
  • Index key facial features like position of nose, eyes and mouth
  • Also needed to normalize the holistic face
Face Recognition

- Holistic methods
  - Face as a whole is indexed (mostly PCA, then classification)

- Feature based
  - Eyes, mouth, nose, etc.

- Hybrid methods
  - Combination of both
Content Based Image Retrieval

- Motivation & Semantic Gap
- Perception
- Color Based Features
- Texture Based Features
- Segmentation
- Object Recognition
- Vector Images
- Evaluation
Vector Images

- In contrast to raster images
  - Description of boundaries, regions & effects
  - Raster is rendered at view time
  - Aspects: Size, scaling, modification

- Applications
  - Graphs & charts
  - Cliparts & illustrations
  - Logos & Fonts
  - 3D models and scenes
Vector Image Retrieval

- Same aspects as for raster images
  - Color, Shape & Texture
- Selection of aspects depends on domain
- Extraction might be easier
  - Shape, edge and texture features
  - Objects are not necessarily defined
Example: Shock Graphs

- Actual shape is reduced to a graph (tree)
  - Mostly preserving a characteristic
- Graphs are indexed
  - Using invariants
  - Low dimensional vector
- Retrieval is done by
  - Spatial access
Example: Graph Isomorphism

- Clipart is reduced to graph
  - Describing color & relation between regions
    - includes, is-neighbour
Example: Graph Isomorphism

● Graphs are compared pairwise
  • Based on the similarity of nodes (colors)
  • And structure (edges and types)
● Corpus is clustered hierarchically
  • Generates a “tree of images”
  • That’s called “metric index”
● Retrieval is based on pairwise check with cluster representatives
  • Efficiency depends on the index.
Content Based Image Retrieval

- Motivation & Semantic Gap
- Perception
- Color Based Features
- Texture Based Features
- Segmentation
- Object Recognition
- Vector Images
- Evaluation
Evaluation

- Evaluation methods are similar to information retrieval tasks:
  - Comparison of methods based on precision & recall
  - Evaluation of application user centered (subjective evaluation)

- For classification tasks
  - False/True Posititves/Negatives
User Centered Evaluation

- Sample Evaluation on the use of content based organization of images
  - User were presented layouts of images
  - Compare the 3 types and a list representation
User Centered Evaluation

- Results:
  - Content based method good for “graphical tasks”
  - Metadata based method depends on annotation
  - Random organization also helps user
    - Identification of “strong” images
    - Stick out of surroundings
Exercise 06

- Rank all 7 images according to your subjective quality rating.
- Assume image 1 is the query and all other images are results
  - Rank the 6 result images according to their similarity to the query image
  - Use your own subjective similarity rating.
Exercise 06

- You will also find the images on the course homepage
- Send me an Excel / Calc / PDF until next lesson.
Thank you ...

... for your attention