Recap: RGB Color Space

- Primaries are monochromatic lights (for monitors, they correspond to the three types of phosphors)
- **Subtractive matching** required for some wavelengths

Recap: Color Perception

- Rods and cones act as filters on the spectrum
  - To get the output of a filter, multiply its response curve by the spectrum, integrate over all wavelengths
  - Each cone yields one number
- Q: How can we represent an entire spectrum with 3 numbers?
  - A: We can’t! Most of the information is lost.
  - As a result, two different spectra may appear indistinguishable. Such spectra are known as metamer.

Recap: Color Sensing

- Such spectra are known as Human Luminance Sensitivity Function

Announcements

- Exercise sheet 3 will be made available this afternoon
  - Histogram based object recognition [today’s topic]
  - Mean-shift segmentation [Thursday’s topic]
  - The exercise will be next Tuesday.
  - Submit your results by Monday night.

- Demo competition
  - Design your own Computer Vision demo! (Based on the techniques from the lecture...)
  - Teams of up to 3 students
  - Demo day after the end of the semester
  - Will send around a poll for a suitable date...

Course Outline

- Image Processing Basics
  - Image Formation
  - Binary Image Processing
  - Linear Filters
  - Edge & Structure Extraction
  - Color
- Recognition
  - Global Representations
- Segmentation
- Local Features & Matching
- Object Recognition and Categorization
- 3D Reconstruction
- Motion and Tracking
Recap: CIE XYZ Color Space

- Established in 1931 by the International Commission on Illumination
- Primaries are imaginary, but matching functions are everywhere positive
- 2D visualization: draw \((x, y)\), where \(x = X/(X+Y+Z)\), \(y = Y/(X+Y+Z)\)

Recap: HSV Color Space

- Hue, Saturation, Value (Brightness)
- Nonlinear - reflects topology of colors by coding hue as an angle.
- Matlab: `hsv2rgb`, `rgb2hsv`.

Color as Low-Level Cue

- Color histograms: Use distribution of colors to describe image
- No spatial information - invariant to translation, rotation, scale
- This lecture will explain how this can be done...

Topics of This Lecture

- Object Recognition
  - Appearance-based recognition
  - Global representations
  - Color histograms
- Recognition using histograms
  - Histogram comparison measures
  - Histogram backprojection
  - Multidimensional histograms
- Probabilistic interpretation
  - Probability density estimation
  - Recognition from local samples
  - Extension: recognition of multiple objects in an image
  - Extension: colored derivatives

Object Recognition
Challenges

- Viewpoint changes
- Translation
- Image-plane rotation
- Scale changes
- Out-of-plane rotation
- Illumination
- Noise
- Clutter
- Occlusion

Appearance-Based Recognition

- Basic assumption
  - Objects can be represented by a set of images (“appearances”).
  - For recognition, it is sufficient to just compare the 2D appearances.
  - No 3D model is needed.

⇒ Fundamental paradigm shift in the 90’s

Global Representation

- Idea
  - Represent each object (view) by a global descriptor.
  - For recognizing objects, just match the descriptors.
  - Some modes of variation are built into the descriptor, the others have to be incorporated in the training data.
    - e.g. a descriptor can be made invariant to image-plane rotations.
  - Other variations:
    - Viewpoint changes
    - Translation
    - Scale changes
    - Out-of-plane rotation
    - Illumination
    - Noise
    - Clutter
    - Occlusion

Color: Use for Recognition

- Color:
  - Color stays constant under geometric transformations
  - Local feature
    - Color is defined for each pixel
    - Robust to partial occlusion

- Idea
  - Directly use object colors for recognition
  - Better: use statistics of object colors

Color Histograms

- Color statistics
  - Here: RGB as an example
  - Given: tristimulus R,G,B for each pixel
  - Compute 3D histogram
    - \( H(R,G,B) = \# \text{pixels with color (R,G,B)} \)

Color Normalization

- One component of the 3D color space is intensity
  - If a color vector is multiplied by a scalar, the intensity changes, but not the color itself.
  - This means colors can be normalized by the intensity.
  - Intensity is given by \( I = R + G + B \):
  - "Chromatic representation"
    - \( r = \frac{R}{R + G + B} \)
    - \( g = \frac{G}{R + G + B} \)
    - \( b = \frac{B}{R + G + B} \)
**Color Normalization**

- Observation:
  - Since \( r + g + b = 1 \), only 2 parameters are necessary
  - E.g., one can use \( r \) and \( g \)
  - and obtains \( b = 1 - r - g \)

**Color Histograms**

- Robust representation

**Topics of This Lecture**

- **Object Recognition**
  - Appearance-based recognition
  - Color histograms

- **Recognition using histograms**
  - Histogram comparison measures
  - Histogram backprojection
  - Multidimensional histograms

- **Probabilistic Interpretation**
  - Probability density estimation
  - Recognition from local samples
  - Extension: recognition of multiple objects in an image
  - Extension: colored derivatives

**Recognition Using Histograms**

- Histogram comparison

- With multiple training views
What Is a Good Comparison Measure?

• How to define matching cost?

Comparison Measures: Kullback-Leibler

– B. Leibe

Comparison Measures: Mahalanobis Distance

– B. Leibe

Comparison Measures: Chi-Square

– B. Leibe

Comparison Measures: Bhattacharyya Distance

– B. Leibe

Comparison Measures: Kullback-Leibler

– B. Leibe
Comp. Measures: Histogram Intersection

- **Definition**
  - Intersection
  \[ \cap(Q, V) = \sum_i \min(q_i, v_i) \]

- **Motivation**
  - Measures the common part of both histograms
  - Range: [0, 1]
  - For unnormalized histograms, use the following formula
  \[ \cap(Q, V) = \frac{1}{2} \left( \frac{\sum_i \min(q_i, v_i)}{\sum_i q_i} + \frac{\sum_i \min(q_i, v_i)}{\sum_i v_i} \right) \]

Comp. Measures: Earth Movers Distance

- **Motivation:** Moving Earth

\[ \text{(distance moved) \times (amount moved)} \]

\[ \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} \times (\text{amount moved}) \]

- **Linear Programming Problem**

\[ Q \]

\[ V \]

\[ m \text{ clusters} \]

\[ n \text{ clusters} \]

\[ \text{All movements} \]
**Comp. Measures: Earth Movers Distance**

- **Motivation: Moving Earth**
  - Linear Programming Problem

\[ \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij} = \text{WORK} \]

- **Constraints**
  1. Move "earth" only from Q to V
  2. Cannot send more "earth" than there is
  3. V cannot receive more than it can hold
  4. As much "earth" as possible must be moved.
    - Either Q must be completely spent or V must be completely filled.

\[ \sum_{j=1}^{n} f_{ij} \leq w_{qj} \]

\[ \sum_{i=1}^{m} f_{ij} \leq w_{vj} \]

**EMD Computation**

- **Motivation: Moving Earth**
  - Linear Programming Problem
  - Distance measure
  \[ D_{EMD} (Q, V) = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij} \]

\[ \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij} = \text{WORK} \]

- **Advantages**
  - Nearness measure without quantization
  - Partial matching
  - A true metric
  - Disadvantage: expensive computation
  - Efficient algorithms available for 1D
  - Approximations for higher dimensions...
Summary: Comparison Measures

- Vector space interpretation
  - Euclidean distance
  - Mahalanobis distance
- Statistical motivation
  - Chi-square
  - Bhattacharyya
- Information-theoretic motivation
  - Kullback-Leibler divergence, Jeffreys divergence
- Histogram motivation
  - Histogram intersection
- Ground distance
  - Earth Movers Distance (EMD)

Comparison for Image Retrieval

- L2 distance
- Jeffrey divergence
- \( \chi^2 \) statistics
- Earth Movers Distance

Histogram Comparison

- Which measure is best?
  - Depends on the application...
  - Euclidean distance is often not robust enough.
  - Both Intersection and \( \chi^2 \) give good performance for histograms.
  - \( \chi^2 \) is a bit more discriminative.
  - Kullback-Leibler works sometimes very well, but is expensive.
  - EMD is most powerful, but also quite expensive
  - There exist many other measures not mentioned here
  - e.g. statistical tests: Kolmogorov-Smirnov, Cramer/Von-Mises

Summary: Recognition Using Histograms

- Simple algorithm
  1. Build a set of histograms \( H = \{ h_i \} \) for each known object
  2. Build a histogram \( h_t \) for the test image.
  3. Compare \( h_t \) to each \( h_i \in H \)
  4. Select the object with the best matching score
     - Or reject the test image if no object is similar enough.

"Nearest-Neighbor" strategy

Topics of This Lecture

- Object Recognition
  - Appearance-based recognition
  - Global representations
  - Color histograms
- Recognition using histograms
  - Histogram comparison measures
  - Histogram backprojection
  - Multidimensional histograms
- Probabilistic interpretation
  - Probability density estimation
  - Recognition from local samples
  - Extension: recognition of multiple objects in an image
  - Extension: colored derivatives

Localization by Histogram Backprojection

- "Where in the image are the colors we're looking for?"
  - Idea: Normalized histogram represents probability distribution
    \[ p(x|\text{obj}) \]
  - Histogram backprojection
    - For each pixel \( x \), compute the likelihood that this pixel color was caused by the object: \( p(x|\text{obj}) \).
    - This value is projected back into the image (i.e. the image values are replaced by the corresponding histogram values).
Color-Based Skin Detection

• Used 18,696 images to build a general color model.
• Histogram representation

M. Jones and J. Rehg, Statistical Color Models with Application to Skin Detection, IJCV 2002.

Localization by Histogram Backprojection

• "Where in the image are the colors we’re looking for?"
  • Query: object with histogram \( M \)
  • Given: image with histogram \( I \)

• Compute the "ratio histogram": 
  \[
  R_i = \min \left( \frac{M_i}{I_{\max}}, 1 \right)
  \]

  - \( R \) reveals how important an object color is, relative to the current image.
  - Color is frequent on the object \( \Rightarrow \) large \( M_i \)
  - Color is frequent in the image \( \Rightarrow \) large \( I \)

  - This value is projected back into the image \( \Rightarrow \) the image values are replaced by the values of \( R \) that they index.
  - The result image is convolved with a circular mask the size of the target object.
  - Peaks in the convolved image indicate detected objects.

Object Localization Results

• Example result after backprojection
  • Looking for blue pullover...


Discussion: Color Histograms

• Pros
  - Invariant to object translation & rotation
  - Slowly changing for out-of-plane rotation
  - No perfect segmentation necessary
  - Histograms change gradually when part of the object is occluded
  - Possible to recognize deformable objects
    - e.g. pullover

• Cons
  - Pixel colors change with the illumination ("color constancy problem")
    - Intensity
    - Spectral composition (illumination color)
  - Not all objects can be identified by their color distribution.

Topics of This Lecture

• Object Recognition
  - Appearance-based recognition
  - Global representations
  - Color histograms

• Recognition using histograms
  - Histogram comparison measures
  - Histogram backprojection
  - Multidimensional histograms

• Probabilistic Interpretation
  - Probability density estimation
  - Recognition from local samples
  - Extension: recognition of multiple objects in an image
  - Extension: colored derivatives

Generalization of the Idea

• Histograms of derivatives
  - \( \mathbf{D}_x \)
  - \( \mathbf{D}_y \)
  - \( \mathbf{D}_{xx} \)
  - \( \mathbf{D}_{xy} \)
  - \( \mathbf{D}_{yy} \)
General Filter Response Histograms

- Any local descriptor (e.g., filter, filter combination) can be used to build a histogram.

- **Examples:**
  - Gradient magnitude: \( \text{Mag} = \sqrt{D_x^2 + D_y^2} \)
  - Gradient direction: \( \text{Dir} = \arctan \frac{D_y}{D_x} \)
  - Laplacian: \( \text{Lap} = D_{xx} + D_{yy} \)

Multidimensional Representations

- Combination of several descriptors
  - Each descriptor is applied to the whole image.
  - Corresponding pixel values are combined into one feature vector.
  - Feature vectors are collected in multidimensional histogram.

Multidimensional Histograms

- **Examples**
  
  ![Feature vector collection](image)

Generalization: Filter Banks

- What filters to put in the bank?
  - Typically we want a combination of scales and orientations, different types of patterns.

  Matlab code available for these examples: [http://www.robots.ox.ac.uk/~vgg/research/texclass/filters.html](http://www.robots.ox.ac.uk/~vgg/research/texclass/filters.html)

Example Application of a Filter Bank

- **Filter bank of 8 filters**
  - 8 response images: magnitude of filtered outputs, per filter
Recall: These looked very similar in terms of their color distributions (when our features were R-G-B).

But how would their texture distributions compare?

Special Case: Multiscale Representations

- Combination of several scales
  - Descriptors are computed at different scales.
  - Each scale captures different information about the object.
  - Size of the support region grows with increasing $\sigma$.
  - Feature vectors capture both local details and larger-scale structures.

Summary: Multidimensional Representations

- **Pros**
  - Work very well for recognition.
  - Usually, simple combinations are sufficient (e.g. $D_x$, $D_y$, Mag-Lap).
  - But multiple scales are very important!
  - Generalization: filter banks

- **Cons**
  - High-dimensional histograms $\Rightarrow$ lots of storage space
  - Global representation $\Rightarrow$ not robust to occlusion

Topics of This Lecture

- **Object Recognition**
  - Appearance-based recognition
  - Global representations
  - Color histograms
  - Recognition using histograms
    - Histogram comparison methods
    - Histogram backprojection
    - Multidimensional histograms

- **Probabilistic Interpretation**
  - Probability density estimation
  - Recognition from local samples
  - Extension: recognition of multiple objects in an image
  - Extension: colored derivatives

From Global To Local...

- Up to now, we have compared entire histograms.

  $\Rightarrow$ Problematic if objects can be partially occluded.

- Now:
  - Look at local measurements only.
  - What can we tell if we only see a single pixel of the object?

Recall: Working with Probabilities

- **Random Variables**:
  - $A, B$

- **Probabilities**:
  - $\Pr(A), \Pr(B)$

- **Joint probability**
  - $\Pr(A, B)$

- **Conditional probability**
  - $\Pr(A \mid B)$
Recall: Manipulation Rules

- Factorization of the joint
  \[ \Pr(A, B) = \Pr(A | B) \Pr(B) = \Pr(B | A) \Pr(A) \]

- Marginalization
  \[ \Pr(A) = \sum_i \Pr(A, b_i) = \sum_i \Pr(A | b_i) \Pr(b_i) \]
  \[ = \sum_i \Pr(b_i | A) \Pr(A) \]

- Bayes theorem
  \[ \Pr(A | B) = \frac{\Pr(B | A) \Pr(A)}{\Pr(B)} \]

Probabilistic Derivation

- Probability of object \( o_n \) given measurement \( m_k \)
  \[ p(o_n | m_k) = \frac{p(m_k | o_n) p(o_n)}{p(m_k)} \]

- Recall: Bayes theorem
  \[ \Pr(A | B) = \frac{\Pr(B | A) \Pr(A)}{\Pr(B)} \]

Probabilistic Recognition

- Assumption: all objects equally probable (“naïve Bayes”)
  \[ p(o_i) = \frac{1}{N} \]
  \[ p(o_n | m_k) = \frac{p(m_k | o_n) p(o_n)}{\sum_i p(m_k | o_i) p(o_i)} \]
  \[ \text{value of hist. cell} \]
  \[ \text{sum over all objects} \]

- Joint probability for two measurements
  \[ p(o_n | m_k \land m_j) = \frac{p(m_k \land m_j | o_n) p(o_n)}{\sum_i p(m_k \land m_j | o_i) p(o_i)} \]
  \[ \text{Assumption: } m_k \text{ and } m_j \text{ are independent} \]
  \[ \text{The individual probabilities can be multiplied} \]
  \[ p(o_n | m_k \land m_j) = \frac{p(m_k | o_n) p(m_j | o_n) p(o_n)}{\sum_i p(m_k | o_i) p(m_j | o_i) p(o_i)} \]
**Probabilistic Recognition**

- Joint probability for $K$ independent measurements
  \[
  p(o_n | m_k) = \frac{\prod_k p(m_k | o_n) p(o_n)}{\sum_k \prod_k p(m_k | o_n) p(o_n)}
  \]

- Assumption: all objects are equally probable
  \[
  p(o_i) = \frac{1}{N}
  \]

  \[
  p(o_n | m_k) = \frac{\prod_k p(m_k | o_n) p(o_n)}{\sum_k \prod_k p(m_k | o_n) p(o_n)}
  \]

**Bayesian Recognition Algorithm**

1. Build up histograms $p(m_k | o_n)$ for each training object.
2. Sample the test image to obtain $m_k, k \in K$.
   - Only small number of local samples necessary.
3. Compute the probabilities for each training object.
   \[
   p(o_n | \text{Image}) = \frac{\prod_k p(m_k | o_n) p(o_n)}{\sum_k \prod_k p(m_k | o_n) p(o_n)}
   \]
4. Select the object with the highest probability
   - Or reject the test image if no object accumulates sufficient probability.

**Practical Issues**

- Most expensive step
  3. Compute the probabilities for each training object.
  
  \[
  p(o_n | \text{Image}) = \frac{\prod_k p(m_k | o_n) p(o_n)}{\sum_k \prod_k p(m_k | o_n) p(o_n)}
  \]

- Notes
  - The numerator computes a score indicating how probable each object $o_i$ in the database is.
  - This score can be used to compare the different object hypotheses.
  - The denominator is the same for all objects in the database.
  - This term is important in order to decide if we have accumulated sufficient evidence to make a decision.

**Results: Probabilistic (Bayesian) Recognition**

- Test database
  - 103 test objects
  - 1327 test images total
  - 607 images with scale changes and rotations for 83 objects
  - 720 images with different viewpoints for 20 objects
  - Use 6D descriptor
    - $D_x$ with $\sigma = \{1,2,4\}$
    - Explicitly trained for scale changes & rotations

- Example image from test database
**Experimental Evaluation**

- Recognition under Partial Occlusion
  - Compare intersection, $\chi^2$, and probabilistic recognition

- Results
  - Intersection more robust to occlusion than $\chi^2$
  - Probabilistic recognition most robust
    - 62% visibility $\Rightarrow$ 100% recognition
    - 33% visibility $\Rightarrow$ 99% recognition
    - 13% visibility $\Rightarrow$ >90% recognition

**Topics of This Lecture**

- Object Recognition
  - Appearance-based recognition
    - Global representation
    - Color histograms
  - Recognition using histograms
    - Histogram comparison measure
    - Histogram backprojection
    - Multi-dimensional histograms

- Probabilistic Interpretation
  - Probability density estimation
  - Recognition from local samples
  - Extension: recognition of multiple objects in an image
  - Extension: colored derivatives

**Extension: Recognition of Multiple Objects**

- Comparison with Hash table
  - $m_i$, $m_j$ vote ($o_n(m_i)$, $o_n(m_j)$)
  - $o_n$($n$) = $\sum$ $o_n(m_i)$

- Probabilistic Recognition
  - $m_i$, $m_j$ vote ($p(o_n(m_i))$, $p(o_n(m_j))$
  - $p(o_n)$($n$) = $\prod$ $\frac{p(o_n(m_i))p(o_n(m_j))}{\sum_{m_i,m_j} p(m_i,m_j)p(o_n(m_i))}$

**Recognition of Multiple Objects**

- Local Appearance Hashing
  - Combination of the probabilistic recognition with a hash table
  - Only relatively small object region is needed for recognition.
  - Divide image into set of (overlapping) regions.
  - Each region votes for a single object.
  - Region votes are combined to vote for the presence of object $n$.

**Recognition Results**

- Test image 1
  - First Match
  - Second Match
  - Third Match

- Test image 2
  - First Match
  - Second Match
  - Third Match
Why Does It Work?

- Histogram Representation
  - Contains no structural description.
  - Many different objects should result in the same histograms.
  ⇒ Why can the approach still distinguish so many objects?

- Explanation
  - Support regions of neighboring descriptors overlap.
  - Neighborhood relations are captured implicitly.

Topics of This Lecture

- Object Recognition
  - Appearance-based recognition
  - Global representations
  - Color histograms

- Recognition using histograms
  - Histogram comparison measures
  - Histogram backprojection
  - Multidimensional histograms

- Probabilistic Interpretation
  - Probability density estimation
  - Recognition from local samples
  - Extension: recognition of multiple objects in an image
  - Extension: colored derivatives

Extension: Colored Derivatives

- YC1C2 color space

- Color-opponent space
  - Inspired by models of the human visual system
  - Y \equiv intensity
  - C1 \equiv red-green
  - C2 \equiv blue-yellow

Application: Brand Identification in Video

Extension: Colored Derivatives

- Generalization: derivatives along
  - Y axis \rightarrow intensity differences
  - C1 axis \rightarrow red-green differences
  - C2 axis \rightarrow blue-yellow differences

- Feature vector is rotated such that \( D_y = 0 \)
  - Rotation-invariant descriptor
Application: Brand Identification in Video

B. Leibe
Perceptual and Sensory Augmented Computing
Computer Vision WS 09/10

Summary

- Appearance-based Object Recognition
  - Using global representations
- Histograms
  - Color histograms
  - Histogram comparison measures
  - Multidimensional histograms
- Probabilistic Recognition
  - Histograms as probability density estimates
  - Recognition from local measurements
  - Recognition of multiple objects in an image

You’re Now Ready for First Applications...

- All the basic components are there
  - Binary processing
  - Filter operators
  - Edges, lines, circles
  - Color
  - Simple global recognition
- So, let’s have some fun!

References and Further Reading

- Background information on histogram-based object recognition can be found in the following paper
- Matlab filterbank code available at
  - http://www.robots.ox.ac.uk/~vgg/research/tesclass/filters.html