Recall: Exercise sheet 3 is due this week
• Hough Transform
• Mean-shift clustering
• Mean-shift segmentation
• Image segmentation with Graph Cuts
  [last Tuesday’s topic]
• The exercise will be on Thursday, 20.11.
  ⇒ Submit your results by Wednesday night.

Course Outline
• Image Processing Basics
• Segmentation
  - Segmentation and Grouping
  - Graph-Theoretic Segmentation
• Recognition
  - Global Representations
  - Subspace representations
• Local Features & Matching
• Object Categorization
• 3D Reconstruction
• Motion and Tracking

Recap: MRFs for Image Segmentation
• MRF formulation
  ⇒ Minimize the energy
  $E(x, y) = \sum_i \phi(x_i, y_i) + \sum_{i,j} \psi(x_i, x_j)$

Recap: Energy Formulation
• Energy function
  $E(x, y) = \sum_i \phi(x_i, y_i) + \sum_{i,j} \psi(x_i, x_j)$

• Unary potentials $\phi$
  - Encode local information about the given pixel/patch
  - How likely is a pixel/patch to belong to a certain class (e.g. foreground/background)?

• Pairwise potentials $\psi$
  - Encode neighborhood information
  - How different is a pixel/patch’s label from that of its neighbor? (e.g. based on intensity/color/texture difference, edges)

Recap: How to Set the Potentials?
• Unary potentials
  - E.g. color model, modeled with a Mixture of Gaussians
  $\phi(x_i, y_i; \theta) = \log \sum_k \theta_k(x_i, k)p(k|x_i)N(y_i; \hat{y}_k, \Sigma_k)$

  ⇒ Learn color distributions for each label
Recap: How to Set the Potentials?

- **Pairwise potentials**
  - **Potts Model**
    
    \[
    \psi(x_i, x_j; \theta_{pq}) = \theta_{pq} \delta(x_i \neq x_j)
    \]
    
    - Simplest discontinuity preserving model.
    - Discontinuities between any pair of labels are penalized equally.
    - Useful when labels are unordered or number of labels is small.
  
  - Extension: "Contrast sensitive Potts model"
    
    \[
    \psi(x_i, x_j, g(y); \theta_{pq}) = -\theta_{pq} g(y) \delta(x_i \neq x_j)
    \]
    
    \[\text{where } g(y) = e^{-\beta |y|} \quad \beta = \frac{1}{2} \left( \text{avg} \left( |y_i| + |y_j| \right) \right)^{-1}\]
    
    - Discourages label changes except in places where there is also a large change in the observations.

Recap: Graph-Cuts Energy Minimization

- Solve an equivalent graph cut problem
  1. Introduce extra nodes: source and sink
  2. Weight connections to source/sink (t-links) by \(\phi(x_i = s)\) and \(\phi(x_i = t)\), respectively.
  3. Weight connections between nodes (n-links) by \(\psi(x_i, x_j)\).
  4. Find the minimum cost cut that separates source from sink.
  
  \[\Rightarrow \text{Solution is equivalent to minimum of the energy.}\]
  
- \(s\)-t MinCut can be solved efficiently
  - Dual to the well-known max flow problem
  - Very efficient algorithms available for regular grid graphs (1-2 MPixels/s)
  - Globally optimal result for 2-class problems

Recap: When Can \(s\)-t Graph Cuts Be Applied?

- \(s\)-t graph cuts can only globally minimize binary energies that are **submodular**.
  
  
  \[E(L) \text{ can be minimized by } s\text{-t graph cuts } \iff E(s, t) + E(t, s) \leq E(s, t) + E(t, s)\]
  
  \[\text{Submodularity } \iff \text{convexity}\]
  
- Submodularity is the discrete equivalent to convexity.
  - Implies that every local energy minimum is a global minimum.
  - \(\Rightarrow\) Solution will be globally optimal.

GrabCut: Data Model

- Obtained from interactive user input
  - User marks foreground and background regions with a brush
  - Alternatively, user can specify a bounding box

GrabCut: Coherence Model

- An object is a coherent set of pixels:
  
  \[
  \psi(x, y) = \gamma \sum_{m=1}^{M} \sum_{n=1}^{N} \delta \left[ x_m \neq x_n \right] e^{-\gamma |x_m - x_n|}
  \]
  
  - How to choose \(\gamma\)?

GraphCut Applications: “GrabCut”

- Interactive Image Segmentation [Boykov & Jolly, ICCV’01]
  - Rough region cues sufficient
  - Segmentation boundary can be extracted from edges
  - **Procedure**
    - User marks foreground and background regions with a brush.
    - This is used to create an initial segmentation which can then be corrected by additional brush strokes.
Iterated Graph Cuts

Result

Energy after each iteration

Iterated Graph Cuts

R

Foreground

G

Color model
(Mixture of Gaussians)

GrabCut: Example Results

Foreground & Background

Background

1

2

3

4

This is included in the newest version of MS Office!

Applications: Interactive 3D Segmentation

Topics of This Lecture

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

• Recognition using histograms
  - Histogram comparison measures
  - Histogram backprojection
  - Multidimensional histograms
  - Extension: colored derivatives

Object Recognition

Challenges

• Viewpoint changes
  - Translation
  - Image-plane rotation
  - Scale changes
  - Out-of-plane rotation

• Illumination
• Noise
• Clutter
• Occlusion
Perceptual and Sensory Augmented Computing

**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
      - Noise
      - Scale changes
      - Clutter
      - Out-of-plane rotation
      - 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"
      \[
      r = \frac{R}{R + G + B}, \quad \quad g = \frac{G}{R + G + B}, \quad \quad b = \frac{B}{R + G + B}
      \]

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

Topics of This Lecture
- Object Recognition
  - Appearance-based recognition
  - Global representations
  - Color histograms
- Recognition using histograms
  - Histogram comparison measures
  - Histogram backprojection
  - Multidimensional histograms
  - Extension: colored derivatives

Recognition Using Histograms
- Histogram comparison

What Is a Good Comparison Measure?
- How to define matching cost?
Comparison Measures: Euclidean Distance

- **Definition**
  - Euclidean Distance (=L₂ norm)
  \[ d(Q, V) = \sum_i (q_i - v_i)^2 \]

- **Motivation**
  - Focuses on the differences between the histograms.
  - Interpretation: distance in feature space.
  - Range: [0, \(\infty\)]
  - All cells are weighted equally.
  - Not very robust to outliers!

Comparison Measures: Mahalanobis Distance

- **Definition**
  - Mahalanobis distance (Quadratic Form)
  \[ d(Q, V) = (Q - V)^T \Sigma^{-1} (Q - V) \]
  \[ = \sum_i \sum_j \frac{(q_i - v_i)(q_j - v_j)}{\sigma_{ij}} \]

- **Motivation**
  - Interpretation:
    - Weighted distance in feature space.
    - Compensate for correlated data.
  - Range: [0, \(\infty\)]
  - More robust to certain outliers.

Comparison Measures: Chi-Square

- **Definition**
  - Chi-square
  \[ \chi^2(Q, V) = \sum_i \frac{(q_i - v_i)^2}{q_i + v_i} \]

- **Motivation**
  - Statistical background:
    - Test if two distributions are different
    - Possible to compute a significance score
  - Range: [0, \(\infty\)]
  - Cells are not weighted equally!
  - More robust to outliers than Euclidean distance.
  - If the histograms contain enough observations...

Comparison Measures: Bhattacharyya Distance

- **Definition**
  - Bhattacharyya coefficient
  \[ BC(Q, V) = \sum_i \sqrt{q_i v_i} \]

- **Common distance measure**
  \[ d_{BC}(Q, V) = \sqrt{1 - BC(Q, V)} \]

- **Motivation**
  - Statistical background:
    - BC measures the statistical separability between two distributions.
  - Range: [0, \(\infty\)]
  - (Reason for \(d_{BC}\): triangle inequality)

Comparison 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
  \[ \cap(Q, V) = \frac{1}{2} \left( \sum_i \min(q_i, v_i) - \sum_i q_i + \sum_i v_i \right) \]
Comp. Measures: Earth Movers Distance

- Motivation: Moving Earth

(distance moved) * (amount moved)

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

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

What is the minimum amount of work to convert Q into V?
EMD Computation

- Constraints

1. Move “earth” only from Q to V

\[ f_{ij} \geq 0 \]

2. Cannot send more “earth” than there is

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

3. V cannot receive more than it can hold

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

4. As much “earth” as possible must be moved.
   - Either Q must be completely spent
   - Or V must be completely filled.

\[ \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = \min \left( \sum_{i=1}^{m} w_{ij}, \sum_{j=1}^{n} w_{ij} \right) \]

Comp. Measures: Earth Movers Distance

- Motivation: Moving Earth
  - Linear Programming Problem
  - Distance measure

\[ D_{EMD}(Q,V) = \frac{\sum_{i,j} d_{ij} f_{ij}}{\sum_{i,j} f_{ij}} \]

- 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.
    - Intersection is a bit more robust.
    - $\chi^2$ is a bit more discriminative.
  - KL/Jeffrey 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
     - Using a suitable comparison measure
     - Or reject the test image if no object is similar enough.

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.
**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., a pullover

- **Cons**
  - Pixel colors change with the illumination
    - "color constancy problem"
  - 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
  - Extension: colored derivatives

**Generalization of the Idea**

- **Histograms of derivatives**
  - $D_x$
  - $D_y$
  - $D_{xx}$
  - $D_{xy}$
  - $D_{yy}$

**General Filter Response Histograms**

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

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

**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**
Multidimensional Representations

- Useful simple combinations
  - $D_xD_y$: Rotation-variant
    - Descriptor changes when image is rotated.
    - Useful for recognizing oriented structures (e.g., vertical lines)
  - Mag-Lap: Rotation-invariant
    - Descriptor does not change when image is rotated.
    - Can be used to recognize rotated objects.
    - Less discriminant than rotation-variant descriptor.

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.

Generalization: Filter Banks

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

Example Application of a Filter Bank

- 8 response images: magnitude of filtered outputs, per filter

Extension: Colored Derivatives

- YC1C2 color space
  \[
  \begin{bmatrix}
  Y \\
  C_1 \\
  C_2
  \end{bmatrix} =
  \begin{bmatrix}
  g_r & g_b & g_b \\
  \frac{3g_y - 2}{2} & \frac{3g_y - 2}{2} & 0 \\
  \frac{9g_y - 2}{2} & \frac{9g_y - 2}{2} & \frac{9g_y - 2}{2}
  \end{bmatrix}
  \begin{bmatrix}
  R \\
  G \\
  B
  \end{bmatrix}
  \]

- Color-opponent space
  - Inspired by models of the human visual system
  - $Y$: Intensity
  - $C_1$: Red-green
  - $C_2$: Blue-yellow

Extension: Colored Derivatives

- Generalization: derivatives along
  - $Y$ axis → intensity differences
  - $C_1$ axis → red-green differences
  - $C_2$ axis → blue-yellow differences

- Feature vector is rotated such that $D_y = 0$
  - Rotation-invariant descriptor
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

Application: Brand Identification in Video

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/texclass/filters.html

false detection