Recap: 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

Recap: 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

“Nearest-Neighbor” strategy
Recap: Histogram Backprojection

- "Where in the image are the colors we’re looking for?"
  - Query: object with histogram $I_j$
  - Given: image with histogram $I_i$
- Compute the "ratio histogram": $R_i = \min \left( \frac{M_{ij}}{I_i}, 1 \right)$
  - $R$ reveals how important an object color is, relative to the current image.
  - Project value back into the image (i.e., replace the image values by the values of $R$ that they index).
  - Convolve result image with a circular mask to find the object.

Recap: Bayesian Recognition Algorithm

1. Build up histograms $p(m_j | o_k)$ for each training object.
2. Sample the test image to obtain $m_t$, $k \in K$.
   - Only small number of local samples necessary.
3. Compute the probabilities for each training object.
   - $p(n_i | m_j) = \prod_j p(n_i | m_j)$
   - $p(n_i | m) = \prod_j p(n_i | m_j)$
   - $p(n_i | m) = \prod_j p(n_i | m_j) p(o_k | m_j)$
4. Select the object with the highest probability
   - Or reject the test image if no object accumulates sufficient probability.

Recap: 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.

Recap: Colored Derivatives

- Generalization: derivatives along $Y$ axis $\rightarrow$ intensity differences
  - $C_1$ axis $\rightarrow$ red-green differences
  - $C_2$ axis $\rightarrow$ blue-yellow differences
- Application:
  - Brand identification in video

Demo Competition

- Design your own Computer Vision demo!
  - Based on the techniques from the lecture...
  - Topic is up to you – it should be fun!
  - Teams of up to 3 students
  - Demo day after the end of the semester
    - Will send around a poll for a suitable date...
    - Participation is optional (but it will be fun!)
    - Demos will count for up to 30 extra exercise points
  - (Small) prizes for best teams

If you have questions, we’ll be happy to give advice...
Topics of This Lecture

- Segmentation and grouping
  - Gestalt principles
  - Image segmentation
- Segmentation as clustering
  - k-Means
  - Feature spaces
- Probabilistic clustering
  - Mixture of Gaussians, EM
- Model-free clustering
  - Mean-Shift clustering
- Graph theoretic segmentation
  - Normalized Cuts

Examples of Grouping in Vision

- Determining image regions
- What things should be grouped?
- What cues indicate groups?
- Object-level grouping
- Figure-ground

Similarity

Symmetry

Common Fate

Proximity
Perceptual and Sensory Augmented Computing
Computer Vision WS 08/09

Muller-Lyer Illusion

- Gestalt principle: grouping is key to visual perception.

The Gestalt School

- Grouping is key to visual perception
- Elements in a collection can have properties that result from relationships
  - “The whole is greater than the sum of its parts”

Gestalt Theory

- Gestalt: whole or group
  - Whole is greater than sum of its parts
  - Relationships among parts can yield new properties/features
- Psychologists identified series of factors that predispose set of elements to be grouped (by human visual system)

“I stand at the window and see a house, trees, sky. Theoretically I might say there were 327 brightnesses and nuances of colour. Do I have 327? No. I have sky, house, and trees.”

Max Wertheimer
(1880-1943)

Untersuchungen zur Lehre von der Gestalt,
Psychologische Forschung, Vol. 4, pp. 301-350, 1923
http://psy.ed.asu.edu/~classics/Wertheimer/Forms/forms.htm

Gestalt Factors

- These factors make intuitive sense, but are very difficult to translate into algorithms.

Continuity through Occlusion Cues

Continuity, explanation by occlusion
Continuity through Occlusion Cues

Figure-Ground Discrimination

Image Segmentation
- Goal: identify groups of pixels that go together

The Goals of Segmentation
- Separate image into coherent “objects”
The Goals of Segmentation

- Separate image into coherent “objects”
- Group together similar-looking pixels for efficiency of further processing

Image Segmentation: Toy Example

- These intensities define the three groups.
- We could label every pixel in the image according to which of these primary intensities it is.
  - i.e., segment the image based on the intensity feature.
- What if the image isn’t quite so simple?

Topics of This Lecture

- Segmentation and grouping
  - Gestalt principles
    - Image Segmentation
- Segmentation as clustering
  - k-Means
  - Feature spaces
  - Probabilistic clustering
    - Mixture of Gaussians, EM
  - Model-free clustering
    - Mean-Shift clustering
- Graph theoretic segmentation
  - Normalized Cuts

Now how to determine the three main intensities that define our groups?
We need to cluster.
Clustering

• With this objective, it is a “chicken and egg” problem:
  - If we knew the cluster centers, we could allocate points to groups by assigning each to its closest center.
  - If we knew the group memberships, we could get the centers by computing the mean per group.

K-Means Clustering

• Basic idea: randomly initialize the k cluster centers, and iterate between the two steps we just saw.
  1. Randomly initialize the cluster centers, \( c_1, \ldots, c_k \).
  2. Given cluster centers, determine points in each cluster
    - For each point \( p \), find the closest \( c_i \). Put \( p \) into cluster \( i \).
  3. Given points in each cluster, solve for \( c_i \)
    - Set \( c_i \) to be the mean of points in cluster \( i \).
  4. If \( c_i \) have changed, repeat Step 2.

Properties

- Will always converge to some solution
- Can be a “local minimum”
- Does not always find the global minimum of objective function:
  \[ \sum_{i} \sum_{p \text{ in } \text{cluster } i} ||p - c_i||^2 \]

Segmentation as Clustering

K-Means Clustering

• Java demo:
  [http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html](http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html)

K-Means++

• Can we prevent arbitrarily bad local minima?
  1. Randomly choose first center.
  2. Pick new center with prob. proportional to \( ||p - c_i||^2 \)
    - (Contribution of \( p \) to total error)
  3. Repeat until \( k \) centers.
  - Expected error = \( O(\log k) \) * optimal

Feature Space

• Depending on what we choose as the feature space, we can group pixels in different ways.
• Grouping pixels based on intensity similarity
  
Feature space: intensity value (1D)
Feature Space
- Depending on what we choose as the feature space, we can group pixels in different ways.
- Grouping pixels based on color similarity
- Feature space: color value (3D)

Segmentation as Clustering
- Depending on what we choose as the feature space, we can group pixels in different ways.
- Grouping pixels based on texture similarity
- Feature space: filter bank responses (e.g., 24D)

Smoothing Out Cluster Assignments
- Assigning a cluster label per pixel may yield outliers:
- How can we ensure they are spatially smooth?

K-Means Clustering Results
- K-means clustering based on intensity or color is essentially vector quantization of the image attributes
  - Clusters don’t have to be spatially coherent
- Clustering based on (r,g,b,x,y) values enforces more spatial coherence
Summary K-Means

- **Pros**
  - Simple, fast to compute
  - Converges to local minimum of within-cluster squared error

- **Cons/Issues**
  - Setting k?
  - Sensitive to initial centers
  - Sensitive to outliers
  - Detects spherical clusters only
  - Assuming means can be computed

Topics of This Lecture

- Segmentation and grouping
  - Gestalt principles
  - Image Segmentation

- Segmentation as clustering
  - k-Means
  - Feature spaces

- Probabilistic clustering
  - Mixture of Gaussians, EM

- Model-free clustering
  - Mean-Shift clustering

- Graph theoretic segmentation
  - Normalized Cuts

Probabilistic Clustering

- **Basic questions**
  - What’s the probability that a point $x$ is in cluster $m$?
  - What’s the shape of each cluster?
- K-means doesn’t answer these questions.

- **Basic idea**
  - Instead of treating the data as a bunch of points, assume that they are all generated by sampling a continuous function.
  - This function is called a **generative model**.
  - Defined by a vector of parameters $\theta$

Mixture of Gaussians

- One generative model is a mixture of Gaussians (MoG)
  - $K$ Gaussian blobs with means $\mu_b$, covariance matrices $V_b$, dimension $d$
    - Blob $b$ defined by:
    - $P(x|\mu_b, V_b)$
    - Blob $b$ is selected with probability $\pi_b$
    - The likelihood of observing $x$ is a weighted mixture of Gaussians
    $P(x|\theta) = \sum_{b=1}^{K} \pi_b P(x|\theta_b), \quad \theta = [\pi_1, \ldots, \pi_K, \mu_1, \ldots, \mu_K]$

Expectation Maximization (EM)

- **Goal**
  - Find blob parameters $\theta$ that maximize the likelihood function:
    $P(\text{data}|\theta) = \prod_x P(x|\theta)$
- **Approach:**
  1. **E-step:** given current guess of blobs, compute ownership of each point
  2. **M-step:** given ownership probabilities, update blobs to maximize likelihood function
  3. Repeat until convergence

EM Details

- **E-step**
  - Compute probability that point $x$ is in blob $b$, given current guess of $\theta$
    $P(b|x, \mu_b, V_b) = \frac{\pi_b P(x|\mu_b, V_b)}{\sum_{b=1}^{K} \pi_b P(x|\mu_b, V_b)}$
- **M-step**
  - Compute probability that blob $b$ is selected
    $\pi_b \text{new} = \frac{1}{N} \sum_{b=1}^{K} P(b|x, \mu_b, V_b)$ (v data points)
  - Mean of blob $b$
    $\mu_b \text{new} = \frac{\sum_{b=1}^{K} \sum_{i} x_i P(b|x, \mu_b, V_b)}{\sum_{b=1}^{K} \sum_{i} P(b|x, \mu_b, V_b)}$
  - Covariance of blob $b$
    $V_b \text{new} = \frac{\sum_{b=1}^{K} \sum_{i} (x_i - \mu_b \text{new})(x_i - \mu_b \text{new})^T P(b|x, \mu_b, V_b)}{\sum_{b=1}^{K} \sum_{i} P(b|x, \mu_b, V_b)}$
Applications of EM

- Turns out this is useful for all sorts of problems
  - Any clustering problem
  - Any model estimation problem
  - Missing data problems
  - Finding outliers
  - Segmentation problems
    - Segmentation based on color
    - Segmentation based on motion
    - Foreground/background separation
  - ...

- EM demo

Summary: Mixtures of Gaussians, EM

- Pros
  - Probabilistic interpretation
  - Soft assignments between data points and clusters
  - Generative model, can predict novel data points
  - Relatively compact storage

- Cons
  - Local minima
  - k-means is NP-hard even with k=2
  - Initialization
    - Often a good idea to start with some k-means iterations.
  - Need to choose generative model
  - Numerical problems are often a nuisance

Topics of This Lecture

- Segmentation and grouping
  - Gestalt principles

- Image segmentation

- Segmentation as clustering
  - k-means

- Feature spaces

- Probabilistic clustering
  - Mixture of Gaussians, EM

- Model-free clustering
  - Mean-Shift clustering

- Graph theoretic segmentation
  - Normalized Cuts

Finding Modes in a Histogram

- How many modes are there?
  - Mode = local maximum of the density of a given distribution
  - Easy to see, hard to compute

Mean-Shift Segmentation

- An advanced and versatile technique for clustering-based segmentation

http://www.caip.rutgers.edu/~comanici/MSPAMI/msPamiResults.html


Slide credit: Steve Seitz

B. Leibe
Mean-Shift Algorithm

1. Initialize random seed, and window W
2. Calculate center of gravity (the "mean") of W: \( \sum_{x \in W} x \cdot \rho(x) \)
3. Shift the search window to the mean
4. Repeat Step 2 until convergence

Slide credit: Steve Seitz & B. Sarel
Mean-Shift

- Cluster: all data points in the attraction basin of a mode
- Attraction basin: the region for which all trajectories lead to the same mode

Real Modality Analysis

- Tessellate the space with windows
- Run the procedure in parallel
- Slide by Y. Ukrainitz & B. Sarel

Mean-Shift Clustering/Segmentation

- Find features (color, gradients, texture, etc)
- Initialize windows at individual pixel locations
- Perform mean shift for each window until convergence
- Merge windows that end up near the same “peak” or mode

Slide credit: Svetlana Lazebnik

The blue data points were traversed by the windows towards the mode.
Mean-Shift Segmentation Results

http://www.caip.rutgers.edu/~comanicii/MSPAMI/msPamiResults.html

More Results

Problem: Computational Complexity

• Need to shift many windows.
• Many computations will be redundant.

Speedups: Basin of Attraction

1. Assign all points within radius r of end point to the mode.

Speedups

2. Assign all points within radius r/c of the search path to the mode.
Summary Mean-Shift

- **Pros**
  - General, application-independent tool
  - Model-free, does not assume any prior shape (spherical, elliptical, etc.) on data clusters
  - Just a single parameter (window size $h$)
  - $h$ has a physical meaning (unlike k-means)
  - Finds variable number of modes
  - Robust to outliers

- **Cons**
  - Output depends on window size
  - Window size (bandwidth) selection is not trivial
  - Computationally (relatively) expensive (~2s/image)
  - Does not scale well with dimension of feature space

Segmentation: Caveats

- We’ve looked at bottom-up ways to segment an image into regions, yet finding meaningful segments is intertwined with the recognition problem.
- Often want to avoid making hard decisions too soon
- Difficult to evaluate; when is a segmentation successful?

Generic Clustering

- We have focused on ways to group pixels into image segments based on their appearance
  - Find groups; “quantize” feature space
- In general, we can use clustering techniques to find groups of similar “tokens”, provided we know how to compare the tokens.
  - E.g., segment an image into the types of motions present
  - E.g., segment a video into the types of scenes (shots) present

References and Further Reading

- Background information on segmentation by clustering and on Normalized Cuts can be found in Chapter 14 of
- More on the EM algorithm can be found in Chapter 16.1.2.
- Read Max Wertheimer’s classic thoughts on Gestalt
  - http://psy.ed.asu.edu/~classics/Wertheimer/Forms/forms.htm
- Try the k-means and EM demos at
  - http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html