Computer Vision - Lecture 7

Segmentation and Grouping

12.11.2009

Bastian Leibe
RWTH Aachen
http://www.mmp.rwth-aachen.de

leibe@umic.rwth-aachen.de
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: Color Sensing

- **Electromagnetic spectrum**

Slide credit: Svetlana Lazebnik
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 metamers.

Slide credit: Steve Seitz
Recap: RGB Color Space

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

![RGB matching functions](image)

- Good for devices, but not for perception...

Slide credit: Svetlana Lazebnik
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 
\[
x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}
\]

Slide credit: Svetlana Lazebnik
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

for recognition

for content-based image retrieval (CBIR)


Blobworld system
Carson et al, 1999

Slide credit: Kristen Grauman
Color as Low-Level Cue

• Color histograms: Use distribution of colors to describe image

• No spatial information - invariant to translation, rotation, scale

Slide credit: Kristen Grauman
Recap: Color Sensing

- Electromagnetic spectrum

Human Luminance Sensitivity Function
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 *metamers*.
Recap: RGB Color Space

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

![RGB matching functions](image)

- Good for devices, but not for perception...

Slide credit: Svetlana Lazebnik
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 
  \[ x = X/(X+Y+Z), \quad y = Y/(X+Y+Z) \]

Matching functions
Recap: HSV Color Space

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

Image from mathworks.com
Color as Low-Level Cue

for recognition

for content-based image retrieval (CBIR)


Blobworld system
Carson et al, 1999

Slide credit: Kristen Grauman
Color as Low-Level Cue

- Color histograms: Use distribution of colors to describe image
- No spatial information - invariant to translation, rotation, scale
- Lecture 8 will explain how this can be done...

Slide credit: Kristen Grauman
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

Grouping video frames into shots

Figure-ground

Slide credit: Kristen Grauman

B. Leibe
Similarity
Symmetry

Slide credit: Kristen Grauman
Common Fate

Image credit: Arthus-Bertrand (via F. Durand)

Slide credit: Kristen Grauman
Proximity
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”

Illusory/subjective contours

http://en.wikipedia.org/wiki/Gestalt_psychology

Slide credit: Svetlana Lazebnik

Image source: Steve Lehar
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,
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.

B. Leibe

Image source: Forsyth & Ponce
Continuity through Occlusion Cues
Continuity through Occlusion Cues

Continuity, explanation by occlusion

B. Leibe
Continuity through Occlusion Cues

B. Leibe

Image source: Forsyth & Ponce
Continuity through Occlusion Cues
Figure-Ground Discrimination
The Ultimate Gestalt?
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

“superpixels”

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
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?
• Now how to determine the three main intensities that define our groups?
• We need to cluster.
• Goal: choose three “centers” as the representative intensities, and label every pixel according to which of these centers it is nearest to.

• Best cluster centers are those that minimize SSD between all points and their nearest cluster center $c_i$:

$$\sum_{\text{clusters } i} \sum_{\text{points } p \text{ in cluster } i} ||p - c_i||^2$$
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:
      $$\sum_{\text{clusters } i} \sum_{\text{points } p \text{ in cluster } i} \|p - c_i\|^2$$
Segmentation as Clustering

```matlab
img_as_col = double(im(:));
cluster_membs = kmeans(img_as_col, K);

labelim = zeros(size(im));
for i=1:k
    inds = find(cluster_membs==i);
    meanval = mean(img_as_column(inds));
    labelim(inds) = meanval;
end
```

Slide credit: Kristen Grauman
K-Means Clustering

- Java demo:
  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 $\frac{||p - c_i||^2}{\text{total error}}$ (Contribution of $p$ to total error)
3. Repeat until $k$ centers.

- Expected error = $O(\log k) \times \text{optimal}$

Arthur & Vassilvitskii 2007
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)

Slide credit: Kristen Grauman
**Smoothing Out Cluster Assignments**

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

Slide credit: Kristen Grauman
Segmentation as Clustering

- Depending on what we choose as the feature space, we can group pixels in different ways.

- Grouping pixels based on intensity + position similarity

⇒ Way to encode both similarity and proximity.

Slide credit: Kristen Grauman
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
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

Slide credit: Kristen Grauman
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: $P(x|\mu_b, V_b) = \frac{1}{\sqrt{(2\pi)^d|V_b|}}e^{-\frac{1}{2}(x-\mu_b)^T V_b^{-1} (x-\mu_b)}$
  - Blob $b$ is selected with probability $\alpha_b$
  - The likelihood of observing $x$ is a weighted mixture of Gaussians
    $$P(x|\theta) = \sum_{b=1}^{K} \alpha_b P(x|\theta_b), \quad \theta = [\mu_1, \ldots, \mu_n, V_1, \ldots, V_n]$$

Slide credit: Steve Seitz
**Expectation Maximization (EM)**

- **Goal**
  - Find blob parameters $\theta$ that maximize the likelihood function:
  $$P(data|\theta) = \prod_x P(x|\theta)$$

- **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 $\theta$
  \[
  P(b|x, \mu_b, V_b) = \frac{\alpha_b P(x|\mu_b, V_b)}{\sum_{i=1}^{K} \alpha_i P(x|\mu_i, V_i)}
  \]

- **M-step**
  - Compute probability that blob $b$ is selected
  \[
  \alpha_{b}^{new} = \frac{1}{N} \sum_{i=1}^{N} P(b|x_i, \mu_b, V_b)
  \]
  ($N$ data points)
  - Mean of blob $b$
  \[
  \mu_{b}^{new} = \frac{\sum_{i=1}^{N} x_i P(b|x_i, \mu_b, V_b)}{\sum_{i=1}^{N} P(b|x_i, \mu_b, V_b)}
  \]
  - Covariance of blob $b$
  \[
  V_{b}^{new} = \frac{\sum_{i=1}^{N} (x_i - \mu_{b}^{new})(x_i - \mu_{b}^{new})^T P(b|x_i, \mu_b, V_b)}{\sum_{i=1}^{N} P(b|x_i, \mu_b, V_b)}
  \]
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

Slide credit: Steve Seitz
Segmentation with EM

Original image

EM segmentation results

k=2  k=3  k=4  k=5

Image source: Serge Belongie
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 know number of components
    - Solutions: model selection (AIC, BIC), Dirichlet process mixture
  - 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

Slide credit: Steve Seitz
Mean-Shift Segmentation

- An advanced and versatile technique for clustering-based segmentation

![Segmented "landscape 1"](image1)

![Segmented "landscape 2"](image2)

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


Slide credit: Svetlana Lazebnik
Mean-Shift Algorithm

- **Iterative Mode Search**
  1. Initialize random seed, and window $W$
  2. Calculate center of gravity (the “mean”) of $W$: $\sum_{x \in W} x H(x)$
  3. Shift the search window to the mean
  4. Repeat Step 2 until convergence
Mean-Shift

Region of interest
Center of mass

Mean Shift vector

Slide by Y. Ukrainitz & B. Sarel
Mean-Shift

Region of interest
Center of mass
Mean Shift vector
Mean-Shift
Mean-Shift

Region of interest

Center of mass

Mean Shift vector
Mean-Shift

Region of interest
Center of mass

Mean Shift vector

Slide by Y. Ukrainitz & B. Sarel
Mean-Shift

Region of interest
Center of mass
Mean Shift vector

Perceptual and Sensory Augmented Computing

Computer Vision WS 08/09

Slide by Y. Ukrainitz & B. Sarel
Mean-Shift
Real Modality Analysis

Tessellate the space with windows

Run the procedure in parallel

Slide by Y. Ukrainitz & B. Sarel
The blue data points were traversed by the windows towards the mode.

Slide by Y. Ukrainitz & B. Sarel
Mean-Shift Clustering

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

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

Slide credit: Svetlana Lazebnik
More Results

Slide credit: Svetlana Lazebnik
More Results
Problem: Computational Complexity

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

B. Leibe
1. Assign all points within radius \( r \) of end point to the mode.

B. Leibe
2. Assign all points within radius $r/c$ of the search path to the mode.

B. Leibe
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 [D. Forsyth, J. Ponce, *Computer Vision - A Modern Approach*. Prentice Hall, 2003](http://psy.ed.asu.edu/~classics/Wertheimer/Forms/forms.htm)

- 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](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](http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html)