# Computer Vision - Lecture 8

## Recognition with Global Representations

18.11.2014

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

leibe@vision.rwth-aachen.de



---

## Reminder

- Exercise sheet 3 is due this week
  - Hough Transform
  - Mean-shift clustering
  - Mean-shift segmentation [last Tuesday's topic]
  - Image segmentation with Graph Cuts [last Thursdays'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

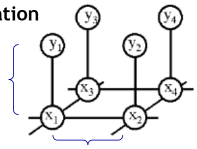


---

## Recap: MRFs for Image Segmentation

- MRF formulation

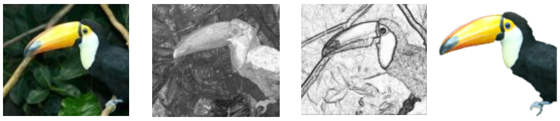
Unary potentials $\phi(x_i, y_i)$



Pairwise potentials $\psi(x_i, x_j)$

⇒ Minimize the energy

$$E(\mathbf{x}, \mathbf{y}) = \sum_i \phi(x_i, y_i) + \sum_{i,j} \psi(x_i, x_j)$$



Data (D) | Unary likelihood | Pair-wise Terms | MAP Solution

Slide adapted from Phil Torr

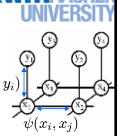


---

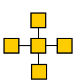
## Recap: Energy Formulation

- Energy function

$$E(\mathbf{x}, \mathbf{y}) = \underbrace{\sum_i \phi(x_i, y_i)}_{\text{Unary potentials}} + \underbrace{\sum_{i,j} \psi(x_i, x_j)}_{\text{Pairwise potentials}}$$




- 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)



B. Leibe



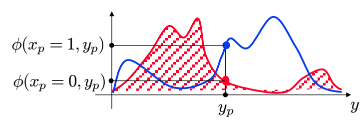
---

## Recap: How to Set the Potentials?

- Unary potentials
  - E.g. color model, modeled with a Mixture of Gaussians

$$\phi(x_i, y_i; \theta_\phi) = \log \sum_k \theta_\phi(x_i, k) p(k|x_i) \mathcal{N}(y_i; \bar{y}_k, \Sigma_k)$$

⇒ Learn color distributions for each label




B. Leibe

## Recap: How to Set the Potentials?



- Pairwise potentials
  - Potts Model
    $$\psi(x_i, x_j; \theta_\psi) = \theta_\psi \delta(x_i \neq x_j)$$
    - 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"
    $$\psi(x_i, x_j, g_{ij}(\mathbf{y}); \theta_\psi) = -\theta_\psi g_{ij}(\mathbf{y}) \delta(x_i \neq x_j)$$
    where
    $$g_{ij}(\mathbf{y}) = e^{-\beta \|y_i - y_j\|^2} \qquad \beta = \frac{1}{2}\left(\text{avg}\left(\|y_i - y_j\|^2\right)\right)^{-1}$$

$\Rightarrow$ Discourages label changes except in places where there is also a large change in the observations.

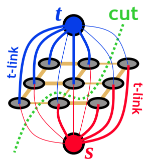


## Recap: Graph-Cuts Energy Minimization

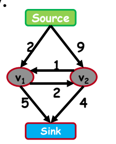


- Solve an equivalent graph cut problem
  1. Introduce extra nodes: source and sink
  2. Weight connections to source/sink (t-links) by $\phi(x_i = s)$ and $\phi(x_i = t)$, respectively.
  3. Weight connections between nodes (n-links) by $\psi(x_i, x_j)$.
  4. Find the minimum cost cut that separates source from sink.

  $\Rightarrow$ 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?



$$E(L) = \underbrace{\sum_p E_p(L_p)}_{\text{t-links}}^{\text{Unary potentials}} + \underbrace{\sum_{pq \in N} E(L_p, L_q)}_{\text{n-links}}^{\text{Pairwise potentials}} \qquad L_p \in \{s, t\}$$

- s-t graph cuts can only globally minimize binary energies that are submodular. [Boros & Hummer, 2002, Kolmogorov & Zabih, 2004]

| *E(L)* can be minimized by *s-t* graph cuts | $\Leftrightarrow$ | $E(s,s)+E(t,t) \leq E(s,t)+E(t,s)$ Submodularity ("convexity") |
|---|---|---|

- Submodularity is the discrete equivalent to convexity.
  - Implies that every local energy minimum is a global minimum.

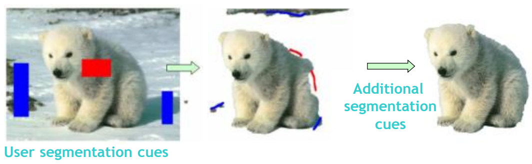
  $\Rightarrow$ Solution will be globally optimal.



## GraphCut Applications: "GrabCut"



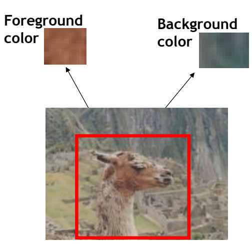
- Interactive Image Segmentation [Boykov & Jolly, ICCV'01]
  - Rough region cues sufficient
  - Segmentation boundary can be extracted from edges
- Procedure
  - User marks foreground and background regions with a brush.
  - This is used to create an initial segmentation which can then be corrected by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






## GrabCut: Data Model






Global optimum of the energy

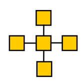
- Obtained from interactive user input
  - User marks foreground and background regions with a brush
  - Alternatively, user can specify a bounding box

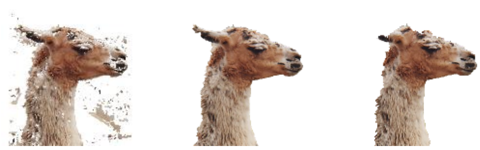


## GrabCut: Coherence Model

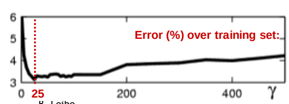


- An object is a coherent set of pixels:
  $$\psi(x, y) = \gamma \sum_{(m,n) \in C} \delta\left[x_n \neq x_m\right] e^{-\beta \|y_m - y_n\|^2}$$



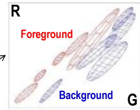


How to choose $\gamma$?

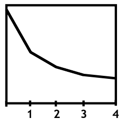


## Iterated Graph Cuts

Result



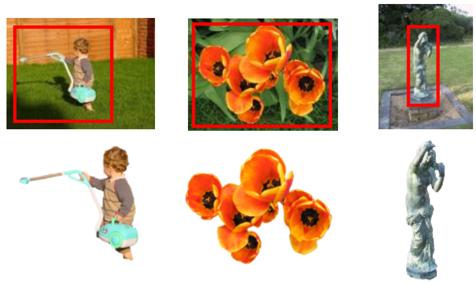

Color model
(Mixture of Gaussians)




Energy after
each iteration



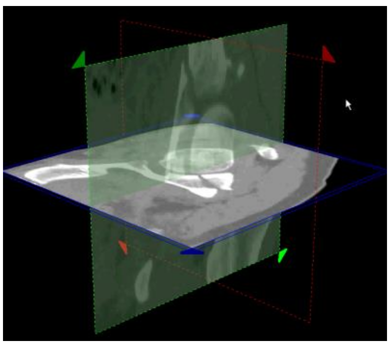
## GrabCut: Example Results



- *This is included in the newest version of MS Office!*


Image source: Carsten Rother

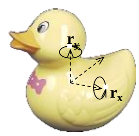
## Applications: Interactive 3D Segmentation




[Y. Boykov, V. Kolmogorov, ICCV'03]

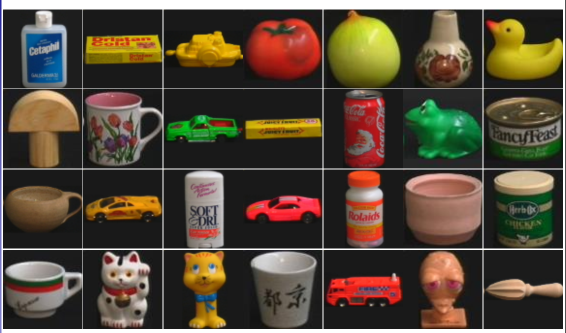
## Topics of This Lecture

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






## Object Recognition

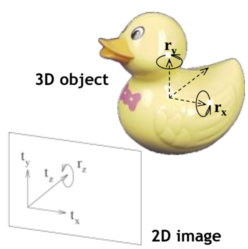




## Challenges

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

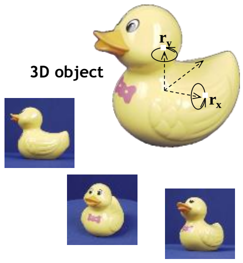
3D object




2D image

## 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.




$\Rightarrow$ 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 | Illumination |
| – Translation | Noise |
| – Scale changes | Clutter |
| – Out-of-plane rotation | Occlusion |



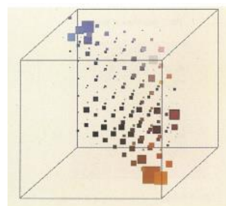
## Color: Use for Recognition

- 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) = #(pixels with color (R,G,B))

[Swain & Ballard, 1991]



## 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"

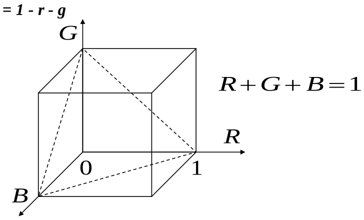
$$r = \frac{R}{R+G+B} \qquad g = \frac{G}{R+G+B}$$

$$b = \frac{B}{R+G+B}$$



## Color Normalization

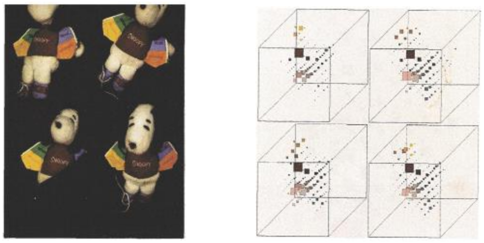
- Observation:
  - Since $r + g + b = 1$, only 2 parameters are necessary
  - E.g. one can use $r$ and $g$
  - and obtains $b = 1 - r - g$



Computer Vision WS 14/15

## Color Histograms

- Robust representation



B. Leibe

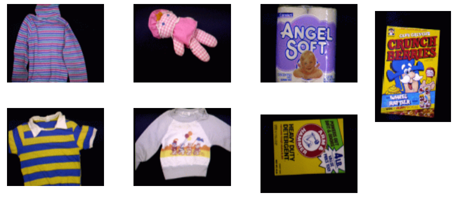
26

[Swain & Ballard, 1991]

Computer Vision WS 14/15

## Color Histograms

- Use for recognition
  - Works surprisingly well
  - In the first paper (1991), 66 objects could be recognized almost without errors




B. Leibe

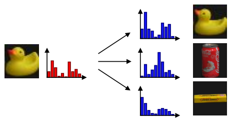
27

[Swain & Ballard, 1991]

Computer Vision WS 14/15

## 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



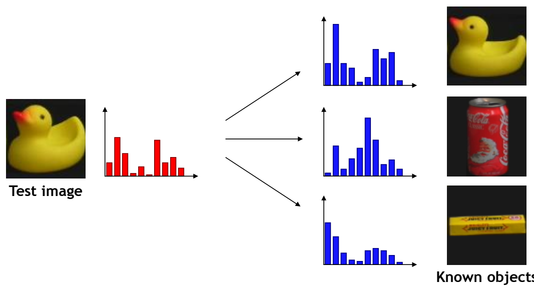
B. Leibe

28

Computer Vision WS 14/15

## Recognition Using Histograms

- Histogram comparison



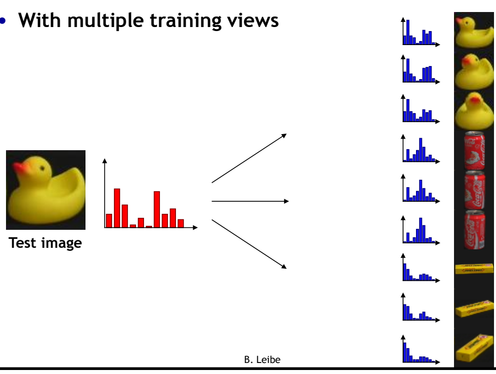

B. Leibe

29

Computer Vision WS 14/15

## Recognition Using Histograms

- With multiple training views



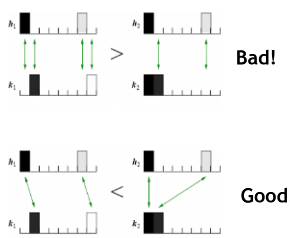

B. Leibe

30

Computer Vision WS 14/15

## What Is a Good Comparison Measure?

- How to define matching cost?



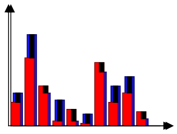

Slide credit: Pete Barnum

31

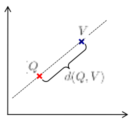
## Comparison Measures: Euclidean Distance

- **Definition**
  - Euclidean Distance (=$L_2$ 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!



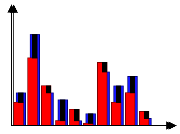




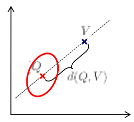

## Comparison Measures: Mahalanobis Distance

- **Definition**
  - Mahalanobis distance(Quadratic Form)

$$d(Q,V) = (Q-V)^\top \Sigma^{-1}(Q-V) = \sum_i \sum_j \frac{(q_i - v_i)(q_j - v_j)}{\sigma_{ij}}$$



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



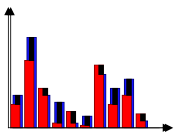





## Comparison Measures: Chi-Square

- **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…





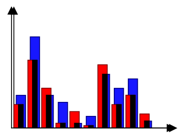
## Comp. Measures: Bhattacharyya Distance

- **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)

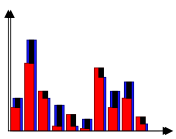




## Comparison Measures: Kullback-Leibler

- **Definition**
  - KL-divergence

$$KL(Q,V) = \sum_i q_i \log \frac{q_i}{v_i}$$



- **Motivation**
  - Information-theoretic background:
    - Measures the expected difference (#bits) required to code samples from distribution $Q$ when using a code based on $Q$ vs. based on $V$.
    - Also called: *information gain, relative entropy*
  - Not symmetric!
  - Symmetric version: *Jeffreys divergence*

$$JD(Q,V) = KL(Q,V) + KL(V,Q)$$

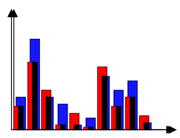




## Comp. Measures: Histogram Intersection

- **Definition**
  - Intersection

$$\cap(Q,V) = \sum_i \min(q_i, v_i)$$

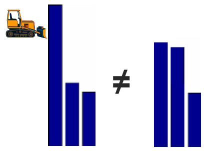


- **Motivation**
  - Measures the common part of both histograms
  - Range: $[0,1]$
  - For unnormalized histograms, use the following formula

$$\cap(Q,V) = \frac{1}{2}\left(\frac{\sum_i \min(q_i,v_i)}{\sum_i q_i} + \frac{\sum_i \min(q_i,v_i)}{\sum_i v_i}\right)$$

## Comp. Measures: Earth Movers Distance

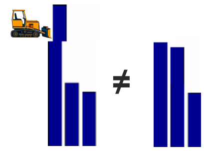
- **Motivation: Moving Earth**






## Comp. Measures: Earth Movers Distance

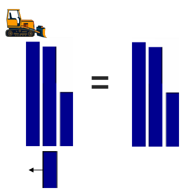
- **Motivation: Moving Earth**






## Comp. Measures: Earth Movers Distance

- **Motivation: Moving Earth**

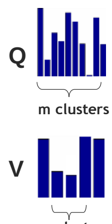


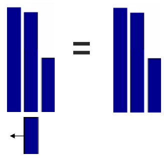

(distance moved) * (amount moved)



## Comp. Measures: Earth Movers Distance

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



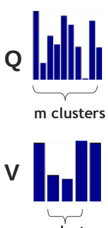

$\sum$ (distance moved) * (amount moved)

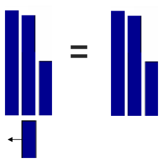
All movements



## Comp. Measures: Earth Movers Distance

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



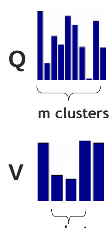

$\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij}$ * (amount moved)

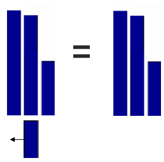
All movements



## Comp. Measures: Earth Movers Distance

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




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

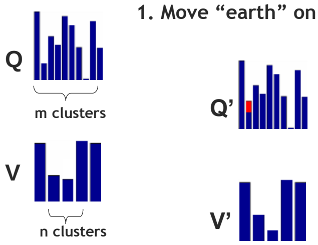
All movements

⇒ ***What is the minimum amount of work to convert Q into V?***

## EMD Computation

- Constraints

1. Move “earth” only from Q to V



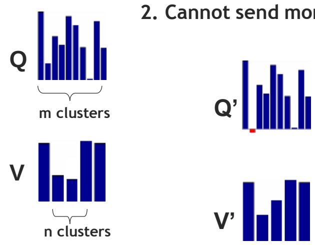

$$f_{ij} \geq 0$$



## EMD Computation

- Constraints

2. Cannot send more “earth” than there is



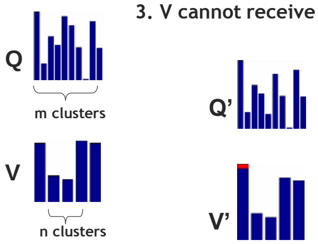

$$\sum_{j=1}^{n} f_{ij} \leq w_{q_i}$$



## EMD Computation

- Constraints

3. V cannot receive more than it can hold



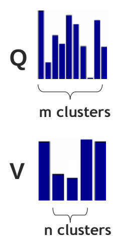

$$\sum_{i=1}^{m} f_{ij} \leq w_{v_j}$$



## EMD Computation

- Constraints

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



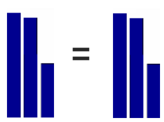

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \min\left(\sum_{i=1}^{m} w_{q_i}, \sum_{j=1}^{n} w_{v_j}\right)$$



## Comp. Measures: Earth Movers Distance

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

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

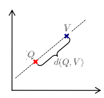



$$\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij} = \mathrm{WORK}$$

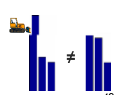
- Advantages
  - Nearness measure without quantization
  - Partial matching
  - A true metric
- Disadvantage: expensive computation
  - Efficient algorithms available for 1D
  - Approximations for higher dimensions…



## 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

Slide credit: Pete Barnum



## 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
     - More exactly, for each *view* of each object
  2. Build a histogram $h_t$ for the test image.
  3. Compare $h_t$ to each $h_i \in H$
     - Using a suitable comparison measure
  4. Select the object with the best matching score
     - Or reject the test image if no object is similar enough.

"Nearest-Neighbor" strategy



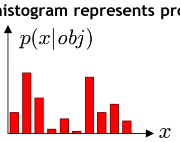
## Topics of This Lecture

- 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

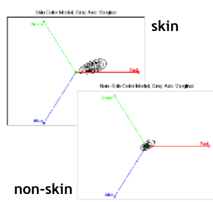



- Histogram backprojection
  - For each pixel $x$, compute the likelihood that this pixel color was caused by the object: $p(x\,|\,obj)$.
  - This value is projected back into the image (*i.e.* the image values are replaced by the corresponding histogram values).



## 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

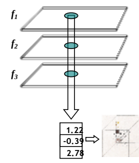
- Pros
  - Invariant to object translation & rotation
  - Slowly changing for out-of-plane rotation
  - No perfect segmentation necessary
  - Histograms change gradually when part of the object is occluded
  - Possible to recognize deformable objects
    - E.g., a pullover
- Cons
  - Pixel colors change with the illumination („color constancy problem")
    - 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
  - Extension: colored derivatives

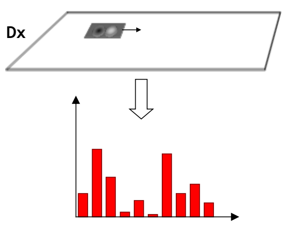







## Generalization of the Idea

- Histograms of derivatives
  - Dx
  - Dy
  - Dxx
  - Dxy
  - Dyy








## General Filter Response Histograms

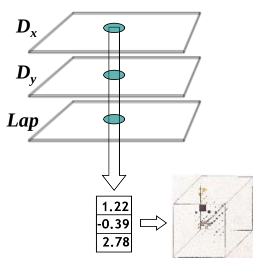
- Any local descriptor (e.g. filter, filter combination) can be used to build a histogram.
- Examples:
  - Gradient magnitude $Mag = \sqrt{D_x^2 + D_y^2}$
  - Gradient direction $Dir = \arctan \frac{D_y}{D_x}$
  - Laplacian $Lap = D_{xx} + D_{yy}$





## 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.

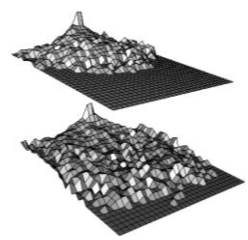







## Multidimensional Histograms

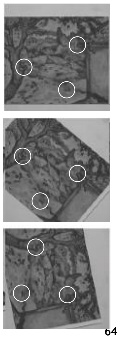
- Examples

[Schiele & Crowley, 2000]

## Multidimensional Representations

- Useful simple combinations
  - $D_x$-$D_y$ — 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.

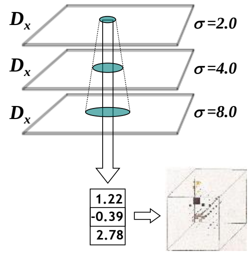






## Special Case: Multiscale Representations

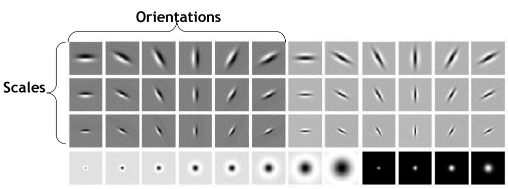
- Combination of several scales
  - Descriptors are computed at different scales.
  - Each scale captures different information about the object.
  - Size of the support region grows with increasing σ.
  - Feature vectors capture both local details and larger-scale structures.








## Generalization: Filter Banks



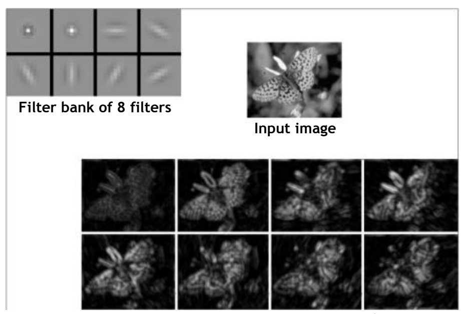

- What filters to put in the bank?
  - Typically we want a combination of scales and orientations, different types of patterns.

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





## Example Application of a Filter Bank



Filter bank of 8 filters

Input image

8 response images: magnitude of filtered outputs, per filter

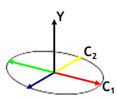




## Extension: Colored Derivatives

- $YC_1C_2$ color space

$$\begin{pmatrix} Y \\ C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} g_r & g_g & g_b \\ \frac{3g_g}{2} & -\frac{3g_r}{2} & 0 \\ \frac{g_b g_r}{g_r^2+g_g^2} & \frac{g_b g_g}{g_r^2+g_g^2} & -1 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

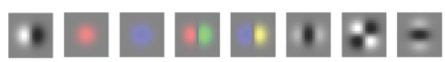



- Color-opponent space
  - Inspired by models of the human visual system
  - Y ≡ intensity
  - $C_1$ ≡ red-green
  - $C_2$ ≡ blue-yellow





## 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 ⇒ lots of storage space
  - Global representation ⇒ not robust to occlusion



## Application: Brand Identification in Video

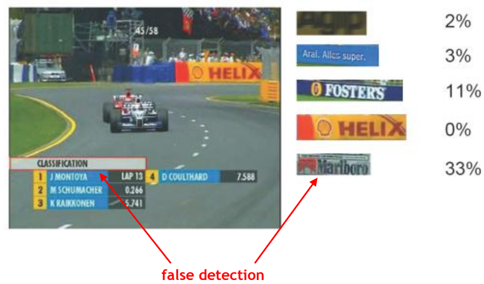

[Hall, Pellison, Riff, Crowley, 2004]

## Application: Brand Identification in Video



[Hall, Pellison, Riff, Crowley, 2004]

## Application: Brand Identification in Video





[Hall, Pellison, Riff, Crowley, 2004]

## References and Further Reading

- Background information on histogram-based object recognition can be found in the following paper
  - B. Schiele, J. Crowley,
    *Recognition without Correspondence using Multidimensional Receptive Field Histograms*.
    International Journal of Computer Vision, Vol. 36(1), 2000.
- Matlab filterbank code available at
  - http://www.robots.ox.ac.uk/~vgg/research/texclass/filters.html