More Histograms

- We’ll again use Histograms as probability distributions, but with some new wrinkles.
  - Instead of comparing two distributions, we want to judge if a single sample comes from a distribution.
  - Want to build a distribution from a histogram with few samples (~100 instead of ~10,000)

Background Subtraction

- Many images of same scene.
- A pixel is foreground or background.
- Many training examples of background.
- Classify pixels in new image
The Problem

Then classify:  

Look at each pixel individually

Then classify:  

Then classify:  

Then classify:  

Just Subtract?

and threshold difference

Background isn’t static
Probability Distribution for Pixels

- \( p(I(x,y)=k) \) for the probability that the pixel at \((x,y)\) will have an intensity of \(k\)

\[
\sum_{k=0}^{255} p(I(x, y) = k) = 1
\]

Bayes’ Law

\[
P(C, D) = P(C \mid D)P(D) \text{ or } P(D \mid C)P(C)
\]

\[
P(C \mid D) = \frac{P(D \mid C)P(C)}{P(D)}
\]

This tells us how to reach a conclusions using evidence, if we know the probability that the evidence would occur.

Probability \((x,y)\) is background if intensity is 107? Who knows?

Probability intensity is 107 if background? We can measure.

\[
P(B(x,y) \mid I(x,y) = k) = \frac{P(I(x,y) = k \mid B(x,y))P(B(x,y))}{P(I(x,y) = k)}
\]
Bayes’ law cont’d

\[
P(B(x, y) | I(x, y) = k) = \frac{P(I(x, y) = k | B(x, y))P(B(x, y))}{P(I(x, y) = k)}
\]

\[
\frac{P(B(x, y) | I(x, y) = k)}{P(F(x, y) | I(x, y) = k)} = \frac{P(I(x, y) = k | B(x, y))P(B(x, y))}{P(I(x, y) = k | F(x, y))P(F(x, y))}
\]

If we have uniform prior for foreground pixel, then key is to find probability distribution for background.

Sample Distribution with Histogram

• **Histogram**: count # times each intensity appears.
• We estimate distribution from experience.
• If 1/100 of the time, background pixel is 17, then assume \( P(I(x,y) = 17 | B) = 1/100 \).
• May not be true, but best estimate.
• Requires *Ergodicity*, ie distribution doesn’t change over time.
Sample Distribution Problems

- This estimate can be noisy.
  Try: \( k=6; \ n=10; \) \texttt{figure(1); hist(floor(k*rand(1,n)), 0:(k-1))}
  for different values of \( k \) and \( n \).
- Need a lot of data.

Histogram of One Pixel Intensities
Kernel Density Estimation

• Assume $p(I(x,y)=k)$ similar for similar values of $k$.
• So observation of $k$ tells us a new observation at **or near** $k$ is more likely.
• Equivalent to smoothing distribution. (smoothing reduces noise)

KDE vs. Sample Distribution

• Suppose we have one observation
  – Sample dist. says that event has prob. 1
    • All other events have prob. 0
  – KDE says there’s a smooth dist. with a peak at that event.
• Many observations just average what happens with one.
KDE cont’d

• To compute $P(x,y)=k$, for every sample we add something based on distance between sample and $k$.
• Let $s_i$ be sample no. $i$ of $(x,y)$, $N$ the number of samples, $\sigma$ be a parameter.

$$P(I(x,y) = k) = \sum_{i=1}^{N} \frac{1}{N\sigma\sqrt{2\pi}} \exp\left(-\frac{(k - s_i)^2}{2\sigma^2}\right)$$

KDE for Background Subtraction

• For each pixel, compute probability background would look like this.
• Then threshold.
Naïve Subtraction

With Model of Background Distribution

Background Subtraction
KDE vs. Binning

• Previously, we use histogram directly to estimate distribution.
• If data is sparse we *bin* data uniformly or with K-means.
  – Eg., divide intensities into 0-15, 16-31, ...
• This is almost like KDE using a box filter.
  – 8 is treated like a uniform distribution from 0-15.
  – But 15 is treated the same way.
• Binning is very fast, but KDE makes more sense.
• The same principle can apply in comparing sparse histograms.
  – Eg, smooth histograms then compare with Chi-squared.