

# Introduction to Image Statistics

Jesus J Caban

## Overview

- Monday, Oct 4<sup>th</sup>: Project Proposal
- Wednesday, Oct 6<sup>th</sup>: Image Relighting (E. Baumel)
- Monday, Oct 11<sup>th</sup>:
  - Camera Shake (F. Zafar)
  - Camera Motion (R. Dighade)
- Wednesday, Oct 13<sup>th</sup>:
  - Assignment #2
  - Color Manipulation (J. Shin)

## Image Statistics & Image Features

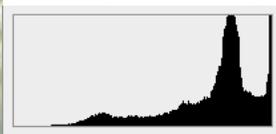
- Estimating statistical properties of an image is important
  - Classification
  - Description
  - Quantification
  - Quality measurement
  - Forensics
  - Reduction
- Statistical properties can be either
  - Global
  - Local



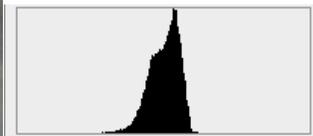
## Image Histogram



Underexposure

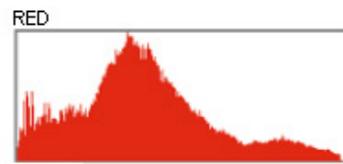
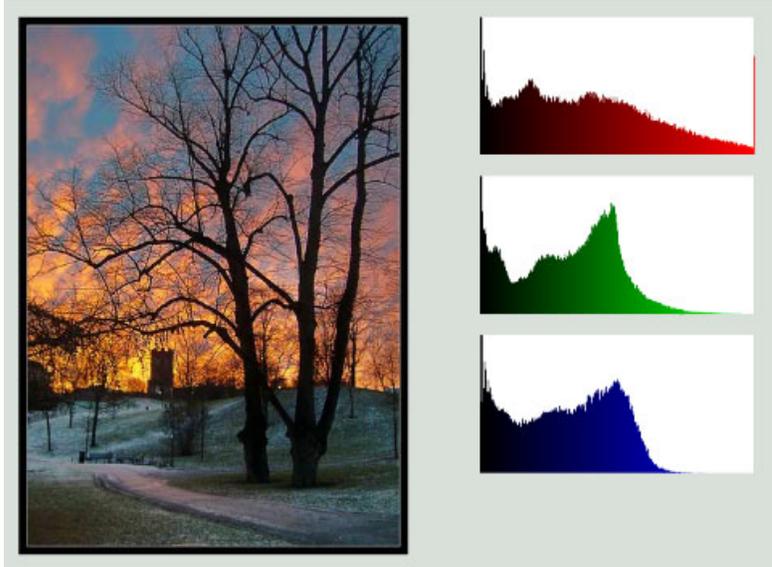


Overexposure



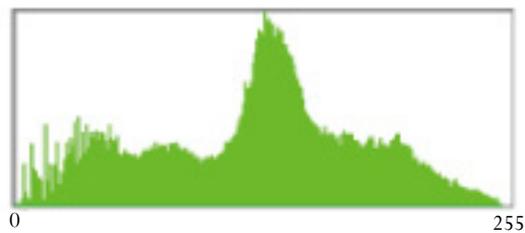
Low contrast

## Color Histograms

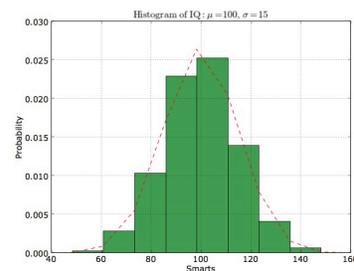
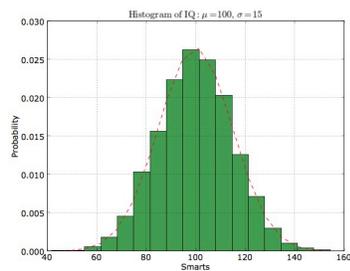
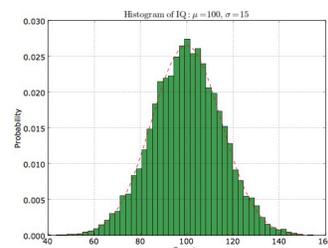
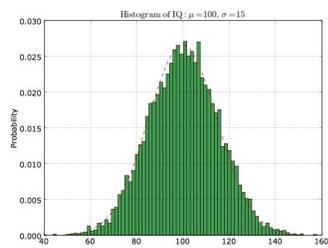


## Histogram

- The likelihood of observing a intensity value at a random location in the image.
- RGB histogram can be done by computing a frequency distribution or histogram for each channel



## Number of bins



## Image Quantization Example

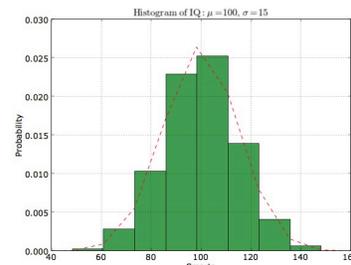
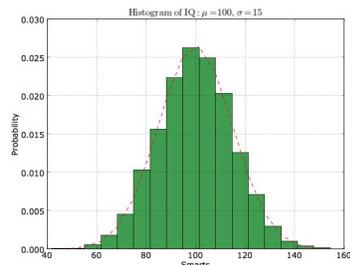
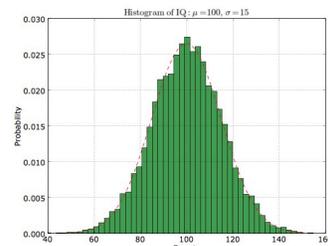
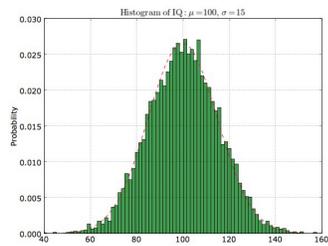
- 256 gray levels (8bits/pixel)    32 gray levels (5 bits/pixel)    16 gray levels (4 bits/pixel)



- 8 gray levels (3 bits/pixel)    4 gray levels (2 bits/pixel)    2 gray levels (1 bit/pixel)



## Number of bins



## Additional Statistical Measurements

1) Mean:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

2) Variance:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

3) Absolute Deviation:

$$\frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

4) Standard Deviation:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

5) Skewness

$$\gamma_1 = \frac{1}{N} \sum_{i=1}^N \left[ \frac{x_i - \mu}{\sigma} \right]^3$$

6) Kurtosis

$$\gamma_2 = \frac{1}{N} \sum_{i=1}^N \left[ \frac{x_i - \mu}{\sigma} \right]^4 - 3$$

## Additional Statistical Measurements

- There are further statistics that describe the shape of the distribution
  - 1<sup>st</sup> moment - Mean (describes central value)
  - 2<sup>nd</sup> moment - Variance (describes dispersion)
  - 3<sup>rd</sup> moment - Skewness (describes asymmetry)
  - 4<sup>th</sup> moment - Kurtosis (describes peakedness)

## Mean

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

- Arithmetic mean (a.k.a. mean or average)
- Describes the central tendency of the histogram
- Not robust given that is often influenced by outliers
- Users can also use median

## Absolute Deviation

$$\frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

- Measures absolute dispersion
- Often implemented as the “Average absolute deviation”

## Variance

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

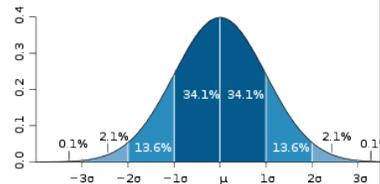
- Describes how far values lie from the mean.
- Also known as one of the moments of a distribution
- 2<sup>nd</sup> standard moment

## Standard Deviation

4) Standard Deviation:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

- Squared root of the variance
- Measures the variability or dispersion of a distribution
- Less robust than the average absolute deviation.



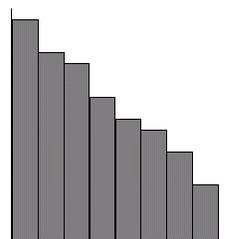
## Skewness

- Skewness measures the degree of asymmetry exhibited by the data

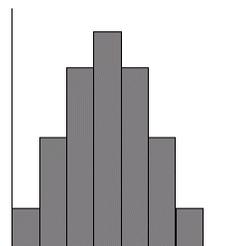
$$\gamma_1 = \frac{1}{N} \sum_{i=1}^N \left[ \frac{x_i - \mu}{\sigma} \right]^3$$

- If skewness equals zero, the histogram is symmetric about the mean
- Positive skewness vs. negative skewness
- 3<sup>rd</sup> standard moment

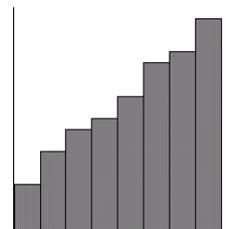
## Skewness



Positively Skewed  
Histogram



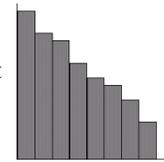
Symmetric Distribution  
Histogram



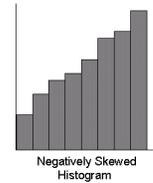
Negatively Skewed  
Histogram

## Skewness

- Positive skewness
  - There are more observations below the mean than above it
  - When the mean is greater than the median



- Negative skewness
  - There are a small number of low observations and a large number of high ones
  - When the median is greater than the mean



## Kurtosis

- Kurtosis measures how peaked the histogram is

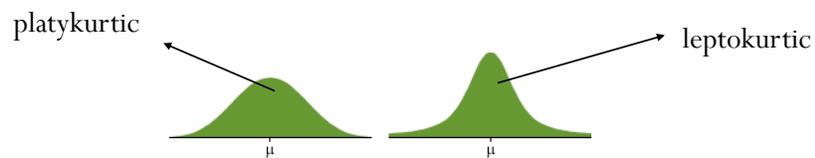
$$\gamma_2 = \frac{1}{N} \sum_{i=1}^N \left[ \frac{x_i - \mu}{\sigma} \right]^4 - 3$$

- The kurtosis of a normal distribution is 0
- Kurtosis characterizes the relative peakedness or flatness of a distribution compared to the normal distribution

## Kurtosis

- Platykurtic:
  - when the kurtosis  $< 0$
  - the frequencies throughout the curve are closer to be equal (i.e., the curve is more flat and wide)
  - relatively flat distribution
- Leptokurtic:
  - When the kurtosis  $> 0$
  - there are high frequencies in only a small part of the curve (i.e., the curve is more peaked)
  - relatively peaked distribution

## Kurtosis



- Kurtosis is based on the size of a distribution's tails.
  - Negative kurtosis (platykurtic) – distributions with short tails
  - Positive kurtosis (leptokurtic) – distributions with relatively long tails

## Why Kurtosis?

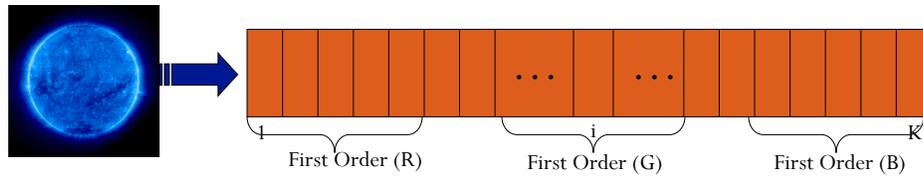


- These two distributions have the same variance, approximately the same skew, but differ in kurtosis.

## Histogram Statistics

- Often called 1<sup>st</sup> order statistic
- Very useful descriptors for an image

## Image Descriptors



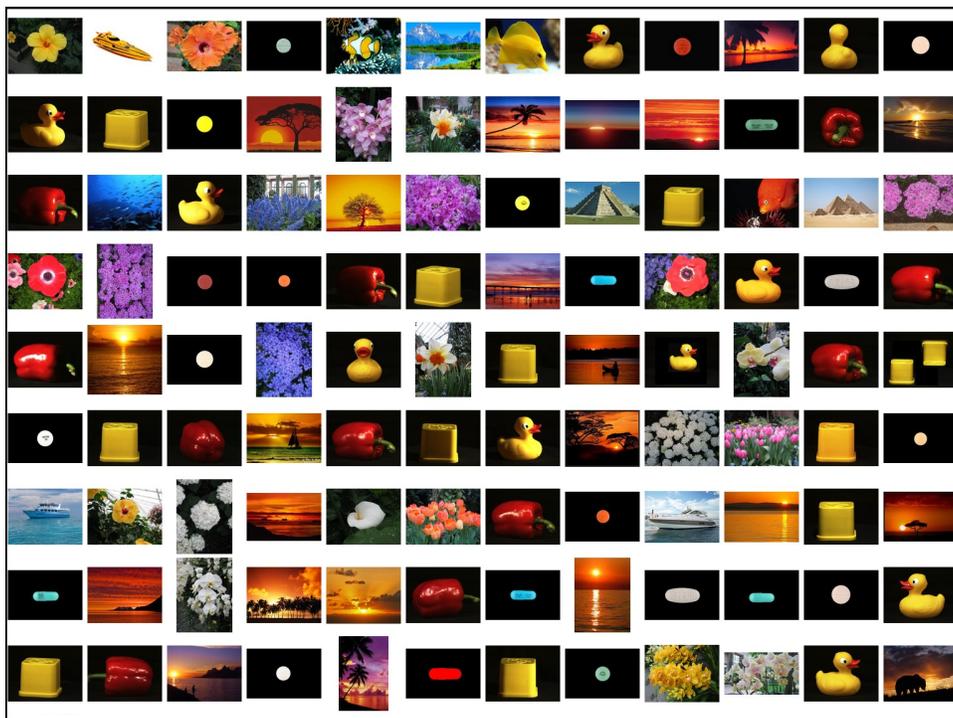
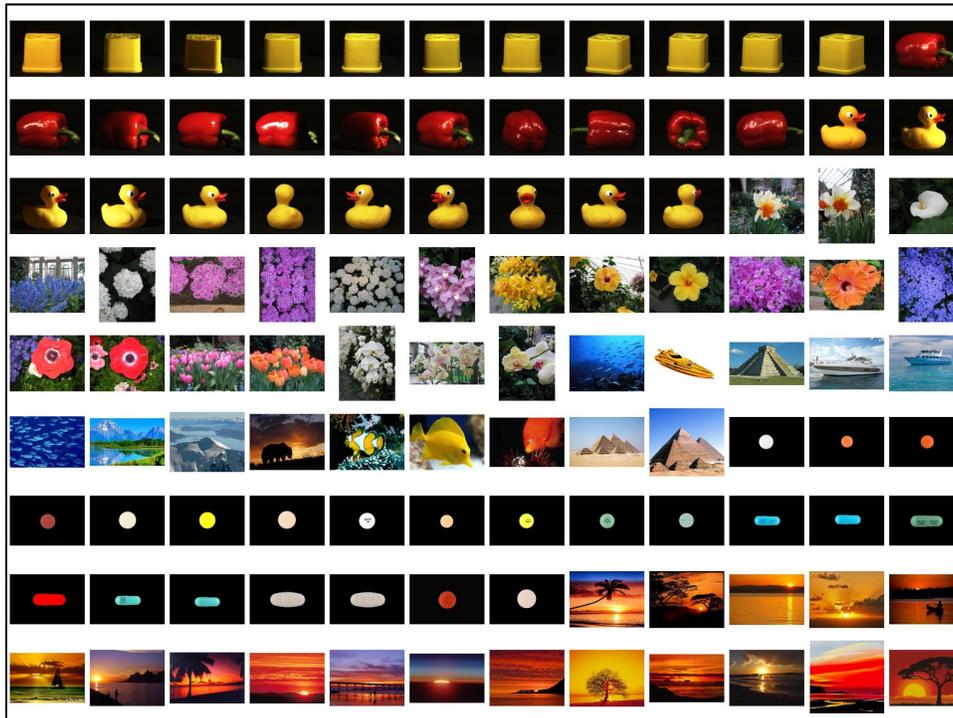
Also known As:

- Image descriptor
- Feature vector
- Features

## Assignment #2

Intro to Image Classification





# 1) Extract Features



Run program:

```
./extract_features image13.jpg
```

Output:

```
image13.jpg,13,0.43,0.2,150,....
```

**What?**

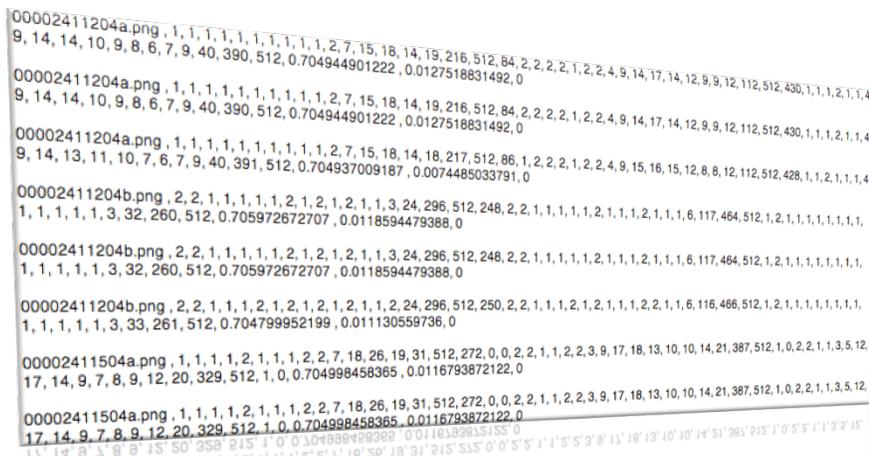
- At least 25 features
- At least capture two different characteristics (color, shape, texture, edge, size, wavelets, etc...)

# 2) CSV file with features

**How?**

In Linux, Mac, Cygwin, etc...

```
$for f in *.jpg; do ./extract_features $f >> all_features.csv; done
```



### 3) Pick 12 closest images

- Given an input image I, pick 12 closest images

**Program:**

```
./closest_images 6 all_features.csv
```

**Output:**

```
[image100.jpg] == image03.jpg image43.jpg ....
```

**How?**

In Linux, Mac, Cygwin, etc... (ImageMagick)

```
$montage image00.jpg image01.jpg .... results1.jpg
```

