Advanced Image Combine Techniques

Ray Gralak

November 16 2008

AIC 2008

Important Equation 1 of 22 (Joke! ☺)

Standard deviation of a discrete random variable or data set

The standard deviation of a discrete random variable is the root-mean-square (RMS) deviation of its values from the mean.

If the random variable X takes on N values $x_1, \ldots, x_N$ (which are real numbers) with equal probability, then its standard deviation σ can be calculated as follows:

1. Find the mean, $\overline{x}$, of the values.
2. For each value x_i calculate its deviation $(x_i - \overline{x})$ from the mean.
3. Calculate the squares of these deviations.
4. Find the mean of the squared deviations. This quantity is the variance σ^2.
5. Take the square root of the variance.

This calculation is described by the following formula:

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

where $\overline{x}$ is the arithmetic mean of the values x_i, defined as:

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i\,.$$

If not all values have equal probability, but the probability of value x_i equals p_i, the standard deviation can be computed by:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} p_i (x_i - \overline{x})^2}{\sum_{i=1}^{N} p_i}} \text{ ,and}$$

$$s = \sqrt{\frac{N' \sum_{i=1}^{N} p_i (x_i - \overline{x})^2}{(N' - 1)\sum_{i=1}^{N} p_i}},$$

where

$$\overline{x} = \frac{\sum_{i=1}^{N} p_i x_i}{\sum_{i=1}^{N} p_i},$$

and N' is the number of non-zero weight elements.

The standard deviation of a data set is the same as that of a discrete random variable that can assume precisely the values from the data set, where the point mass for each value is proportional to its multiplicity in the data set.

Outline

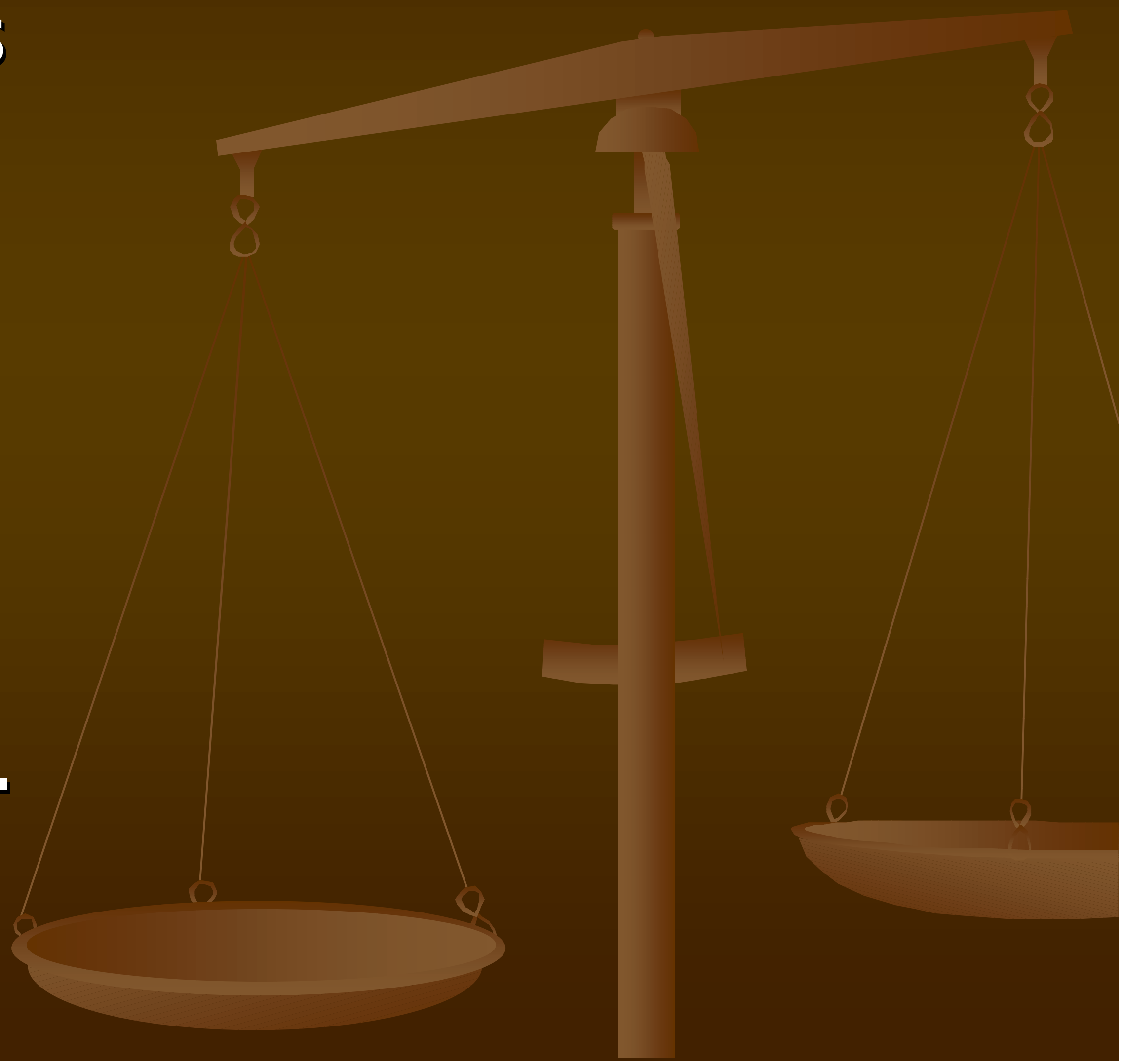
- Some History
- The Imager's "Enemies" – artifacts!
- The Basic Raw Image Types
- Image Calibration Workflow
- Common Combine Algorithm Choices
- Combine Algorithm Details
- Choosing the Right Combine Algorithm
- Some New Algorithm Ideas

History

- Film to CCD transition started in the 90's
- CCDs produced better results with shorter and fewer images
- Stacking CCD images was quickly adopted
- New software (DDP) was invented to show the entire dynamic range
- Result: Unprecedented faint detail BUT with artifacts

The “Enemies”

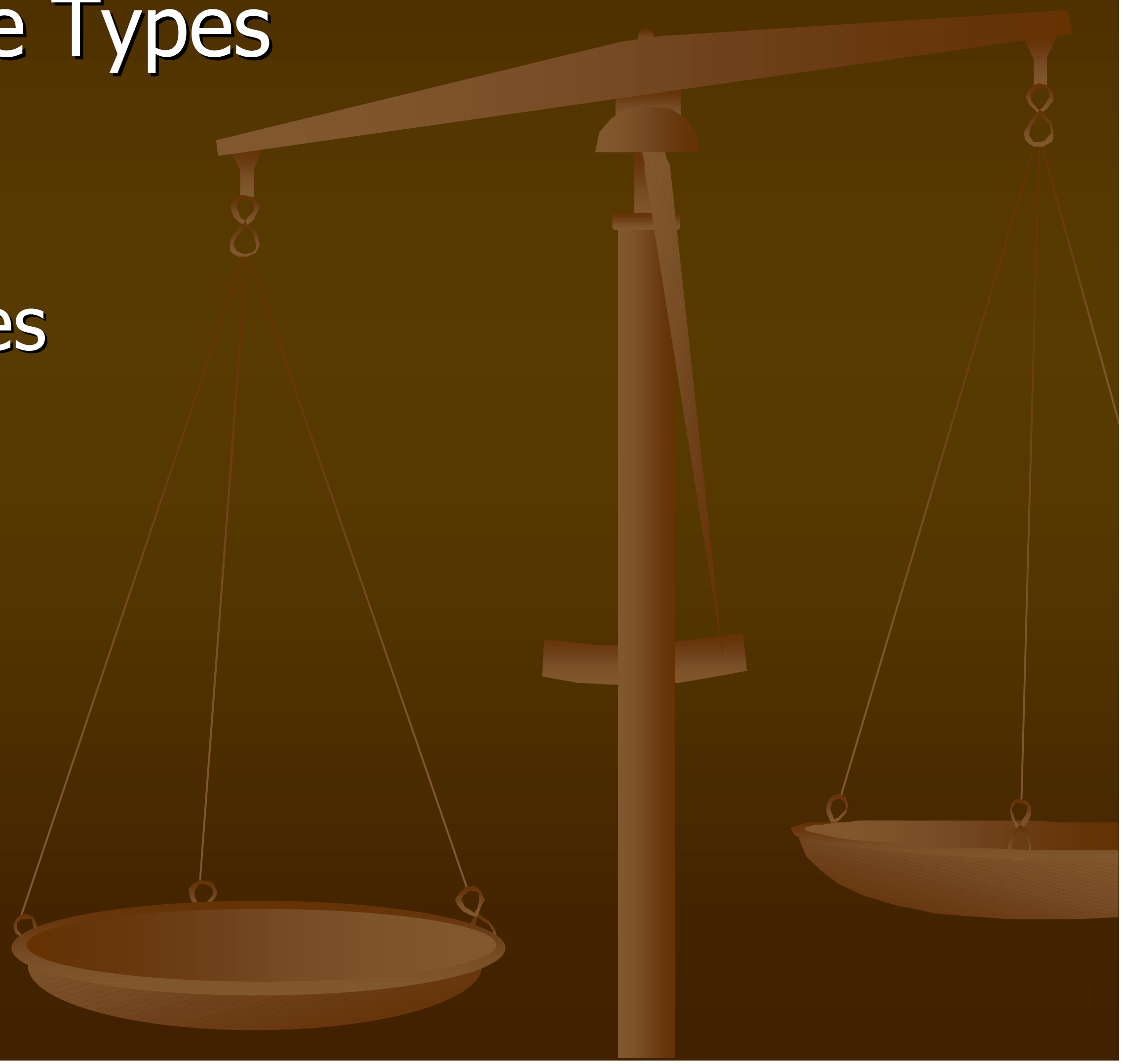
- Cosmic Ray Hits
- Hot/Cold Pixels
- Bad Columns
- Satellite Trails
- Plane Trails
- Asteroids
- Dust Motes, etc.



Some Basics

- Four Raw Image Types
  - Bias Frames
  - Dark Frames
  - Flat Field Frames
  - Light Frames

- Constraints
  - Temperature
  - Time

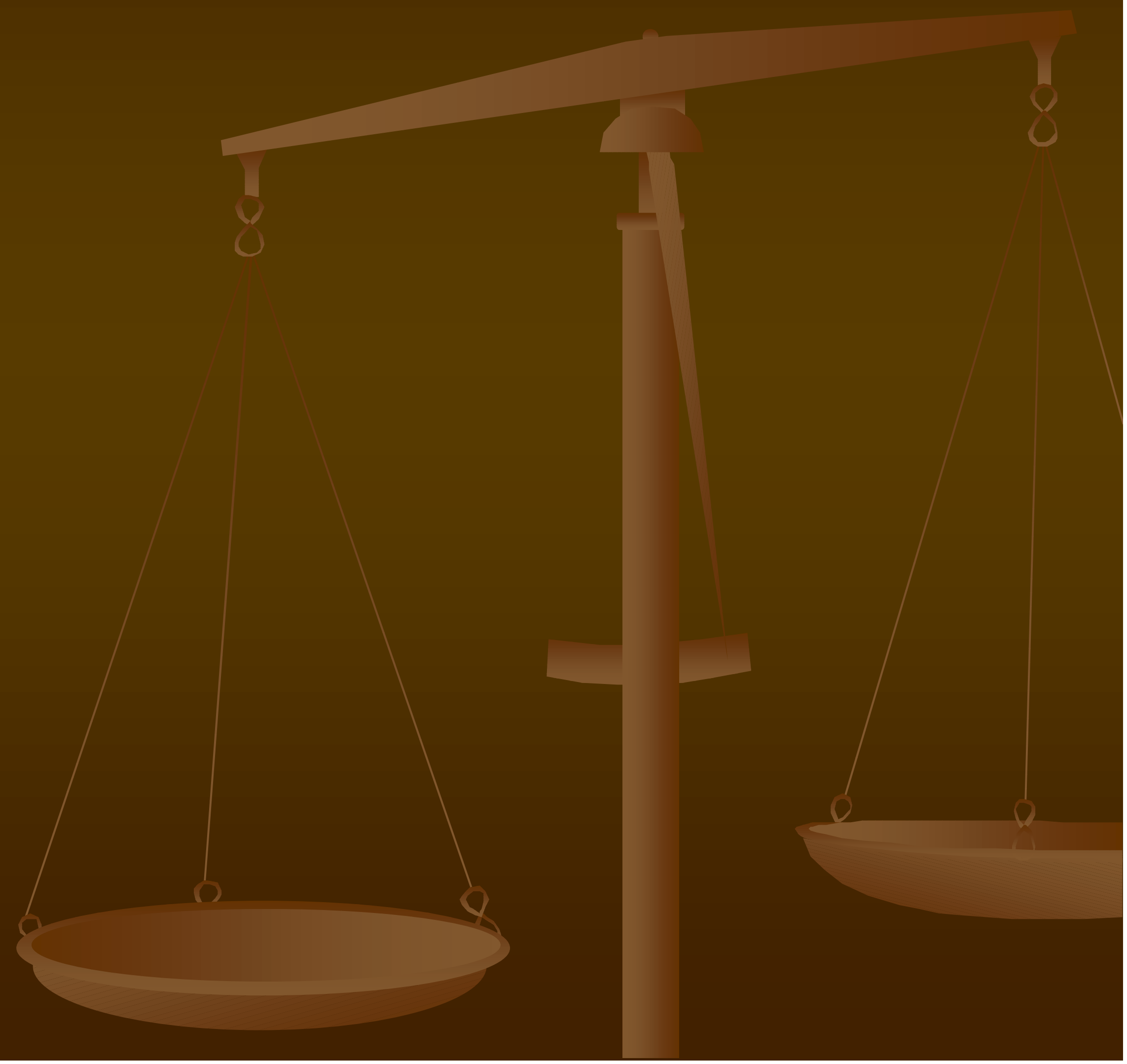


Bias Frames

- Consists of:
  - Bias Level - can vary with time
  - Bias structure - may change slowly over time
  - Read-out Noise

Dark Frames

- Consists of:
  - Bias level
  - Bias structure
  - Read-out noise
  - Dark-current



Flat-Field Frames

- Light frames used to correct variations in:
  - Illumination
  - Pixel to Pixel Sensitivity

- Common ways to create flat-field frames:
  - Light Box
  - Flat target
  - Twilight exposures

Light Frames

- Consist of:
  - Light collected from imaging target/sky
  - Dark Current
  - Bias level, bias structure, and readout noise.

Quick Data Collection Tips

1. Take all frames at the same Temperature.
2. Dark frame count = Light Frame Count
3. Dark Frame duration = Light Frame Duration
4. "Dither" all Light Frames and Flat Frames
5. Flat Frames should be at/near Focus
6. Flat Frames should be at Light Frame Camera Orientation
7. Flat Frames for every filter!

Image Calibration Workflow

1. Combine Bias Frames to Create a Master Bias
2. Combine Dark Frames to Create Master Dark
3. Subtract Master Bias from Each Flat Frame
4. Normalize then Combine Flat Frames
5. Subtract Master Dark from Light Frames
6. Apply Normalized Flat Field to Light Frames
7. Align Light Frames
8. Combine Light Frames

Common Combining Algorithms

- Average
- Median
- Min/Max Clip
- Sigma Clip
- SDM (Standard Deviation Masking)

Average

- Algorithm:
  - Sum of pixels divided by count of frames
- Advantage:
  - Highest Signal/Noise
- Disadvantage:
  - Artifacts are not rejected
- Example:
  - 5,7,12 = (5+7+12)/3 = 24/3 = 8

Median

- Algorithm:
  - Sort pixels from highest to lowest value and pick the one in the middle.
- Advantage:
  - Best noise rejection
- Disadvantage:
  - S/N is lower than Average combine.
- Example:
  - Median of 10,12,**25**,35,54 is 25

Min/Max Clip

- Algorithm
  - Throw away the highest and lowest pixel values and average the rest
- Advantage
  - Rejects most artifacts
- Disadvantages
  - Still can leave some artifacts
  - Works best with 6+ images.
- Example
  - 10,**12,14,18,20**,30 Min/Max Clip = (12+14+18+20)/4 = 16

Sigma Clip

- Algorithm:
  - Calculates the average and standard deviation of the pixels. Averages pixels that are within a Sigma factor range of the average.
- Advantage:
  - Strong noise rejection
- Disadvantages:
  - Needs many images (10+) to work best
  - Bad or failed rejections can occur.

Sigma Clip Example

Suppose the pixel set consists of the values 3, 7, 7, and 19

Step 1: Calculate mean:

(3 + 7 + 7 + 19)/4 = 9

Step 2: Calculate deviation from mean:

3 – 9 = -6

7 – 9 = -2

7 – 9 = -2

19 – 9 = 10

Step 3: Square each deviation

(-6)*(-6) = 36,

(-2)*(-2) = 4

(-2)*(-2) = 4

(10)*(10) = 100

Step 4: Find mean of all deviations

(36 + 4 + 4 + 100) / 4 = 36

Step 5: Calculate Standard Deviation by taking the square root:

SQR(36) = 6

Step 6: S is the user definable Sigma factor. Let S=0.5 in this case

Step 7: Reject values outside of Mean – S * Standard Deviation:

Min = 9 – (0.5 * 6) = 6

Max = 9 + (0.5 * 6) = 12

In values: 3, **7, 7**, and 19:

Reject 3 and 19.

Step 8: Average the remaining values

(7 + 7) / 2 = 7

SDM Example

Suppose we have a stack of 3x3 pixel images

Step 1: calculate the mean, median and standard deviation values :

4	3	1
4	7	8
7	2	1

Mean

3	4	5
6	4	4
2	1	2

Median

3	3	1
4	4	5
13	2	1

Standard Deviation

Step 2: calculate the average of the standard deviations:

$(3+3+1+4+4+5+13+2+1)/9 = 36/9 = 4$

Recomendations

Action	Combine Method
Create Master Bias or Master Dark	SDM: Factor=1.0, Passes=1-3, Normalization=None, Ignore Black Pixels=No, Despeckle=NO
Create Master Flat	1) Median after Normalization 2) SDM: Factor=1.0, Passes=1-3, Normalization=Linear, Area=30%, Ignore Black Pixels=No, Despeckle=Yes
Combine Light Frames	1) SDM: Factor=1.0, Passes=3, Normalization=Linear, Area=30%, Ignore Pixels over 1/3 well capacity, Ignore Black Pixels=Yes, Despeckle=Yes (Sigma factor 1.0)

SDM

- Algorithm:
  - Creates median and average combined images
  - Creates a standard deviation image (mask)
  - Based on the standard deviation at an X,Y pixel either the median or the average pixel is used.
  - Details: http://www.gralak.com/Sigma/SigmaClippingAlgorithmDesignSummary.pdf
- Advantages:
  - Works well with as little as 3 images, very strong noise rejection
  - Not “fooled” by random sets of bad pixels
- Disadvantages
  - Requires more resources (CPU + memory)

Improved SDM

- Mix SDM with other algorithms (e.g. Min/Max Clip, then SDM)
- Standard deviation values vary by level
- Better normalization algorithm (use median of block regions)
- Automatic rejection of pixels surrounding a cluster of bad pixels in a particular image

Summary and Conclusion

- The imager's "Enemies"
- The basic raw image types
- Image calibration workflow
- Common combine algorithm choices
- Combine algorithm details
- Choosing the right combine algorithm
- Some new algorithm ideas