Shack-Hartmann primer &
Correlation wave-front sensing

Lisa A. Poyneer
Lawrence Livermore National Lab

IPAM: Estimation and Control Problems in Adaptive Optics
UCLA, January 22-24, 2004

UCRL-PRES-201939

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under contract No.W-7405-Eng-48.
Goals of this talk

- Explanation of Shack-Hartmann sensors, probabilistic analysis of centroid algorithm
- Detailed performance analysis of correlation sensing using arbitrary images
- Comparison of centroiding to correlation for point-source applications
Why analyze algorithm performance?

- System design may be constrained
  - e.g. read noise requirements for CCD
- Identify important algorithmic/processing issues
  - e.g. do we need smart background subtraction?
- Confirm system performance via rigorous testing of predictions
  - very useful for ‘debugging’ the system
- Possible use in real-time to enhance system performance
Shack-Hartmann sensing measures the phase gradient across a lenslet

Fourier optics principle:
- each lenslet forms a spot on the CCD
- this spot moves with the average slope (linear phase term) in that subaperture

Figure from Gary Chanan
The x-slope is based on the difference between linear combinations of the pixels on the right and left sides.

\[ \hat{x} = \frac{r - l}{t} \]
Linearize formula to get tractable equation for analysis

- Each pixel value is a random variable
- Slope estimate is a non-linear function of the pixels
  - need full probability distributions if quotient is involved
- Linearize function around mean values of each variable using partial derivatives

\[ f(x, y) \approx (x - m_x)f_x(m_x, m_y) + (y - m_y)f_y(m_x, m_y) + f(m_x, m_y) \]
Centroid analysis

- Linearize, then use basic identities to get mean, variance

\[
\hat{x} \approx \frac{r}{m_t} - \frac{l}{m_t} - \frac{t(m_r - m_l)}{m_t^2} + \frac{m_r - m_l}{m_t}
\]

\[
E[\hat{x}] = \frac{m_r - m_l}{m_t}
\]

\[
\text{Var}(\hat{x}) = \frac{\sigma_r^2 + 2m_r^2 + \sigma_l^2 + 2m_l^2 - 4m_r m_l}{m_t^2} + \frac{\sigma_t^2(m_r - m_l)^2}{m_t^4} + \frac{2(m_r - m_l)}{m_t^3} \left( E[lt] - E[rt] \right)
\]
Values of $r$, $l$, $t$ depend on algorithm

- Weight pixels by distance for center
  - $N \times N$ has center at value $N/2 - 1/2$

\[
\begin{align*}
  r &= \sum_{i=N/2}^{N-1} \sum_{j=0}^{N-1} (i - N/2 + 1/2) s[i, j] \\
  l &= \sum_{i=0}^{N/2-1} \sum_{j=0}^{N-1} (-i + N/2 - 1/2) s[i, j] \\
  t &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} s[i, j]
\end{align*}
\]
Probabilistic model for pixel values

- Each pixel in the spot has an normalized expected number of counts (poisson) scaled by exposure time + white read noise

\[ s[i, j] = f \times p[i, j] + n[i, j] \]

- Mean and variance are

\[ m_s[i, j] = f \lambda[i, j] \quad \sigma^2_s[i, j] = f^2 \lambda[i, j] + \sigma_n^2 \]
Centroid variance is made of two independent terms

- Error due to photons is inversely related to SNR

\[ \text{Var}(\hat{x})_p = \frac{\sigma_r^2 + \sigma_l^2}{fm_t^2} \]

RMS error is inverse power law in SNR

- Error variance due to read noise depends on the square of the total number of pixels

\[ \text{Var}(\hat{x})_n = \sigma_n^2 \left( \frac{N^2(N^2 - 1)}{12m_t^2} \right) \]
Background light reduces gain of centroid estimate

- With no background, the expected slope is

\[ E[\hat{x}] = \frac{m_r - m_l}{m_t} \]

- With poisson background (mean \( b \)) added in, expected slope is now

\[ E[\hat{x}] = \frac{m_r - m_l}{m_t + N^2b} \]

Incorrect background subtraction, even by a few counts, can lead to significant loses in gain and increased error.
Enough on centroiding for now...

- Any questions??

- On to correlation of arbitrary images
Why limit AO to scenarios with a point source?

- Almost all AO systems use a point source to measure the phase
- Many interesting scenarios don’t have a point source available
  - observation of Earth from space with light-weight optics w/ time-varying aberrations
  - horizontal and slant-path imaging with small telescope
  - and of course solar astronomy, where AO using small images of solar granulation is successful!
Use a Shack-Hartmann array to form subimages of the scene.

- For pupil-conjugate phase aberration, a subimage of the scene will shift just like a point source.
- Need to field stop down.

WFS camera image of one quadrant of subapertures.

Camera image from C. Thompson and R. Sawvel.
How to best find image shift?

- Find shift between reference $r[m,n]$ and subimage $s[m,n]$
- Rich field, many possible options including
  - non-random parameter estimation
  - deconvolution (linear phase fitting)
  - correlation implementation of MMSE metric

$$\tilde{S}[k, l] = \tilde{R}[k, l] \exp \left[ \frac{-j2\pi(x_0k + y_0l)}{N} \right]$$

Best solution was ‘aliased’ correlation which is fastest way to get energy-normalization in this case
Estimate shift with a correlation-based algorithm

\[ C[m, n] = \sum_i \sum_j r[i - m, j - n]s[i, j] \]

\[ \hat{x}_0 = \Delta x + \frac{0.5(C_{-1} - C_1)}{C_{-1} + C_1 - 2C_0} \]

Parabolic interpolation
Statistical analysis of correlation formula

\[ \hat{x}_0 \approx \Delta_x + \left[ C_{-1}(m_1 - m_0) + C_0(m_{-1} - m_1) + C_1(m_0 - m_{-1}) 
+ 0.5(m_{-1} - m_1)(m_{-1} + m_1 - 2m_0) \right] \times (m_{-1} + m_1 - 2m_0)^{-2} \]

\[ E[\hat{x}_0] = \Delta_x + \frac{0.5(m_1 - m_{-1})}{m_{-1} + m_1 - 2m_0} \]

\[ \sigma_x^2 = [\sigma_{-1}^2(m_1 - m_0)^2 + \sigma_0^2(m_{-1} - m_1)^2 + \sigma_1^2(m_0 - m_{-1})^2 
+ 2(m_1 - m_0)(m_{-1} - m_1)\sigma_{-1,0}^2 + 2(m_1 - m_0)(m_0 - m_{-1})\sigma_{-1,1}^2 
+ 2(m_{-1} - m_1)(m_0 - m_{-1})\sigma_{0,1}^2] \times (m_{-1} + m_1 - 2m_0)^{-4} \]
Use same statistical model for pixels

- Express means, variances of correlation points in terms of pixel models

\[
\begin{align*}
\mathbb{E}[C_k] &= m_k = \sum_i \sum_j m_r[i - k, j]m_s[i, j] \\
\text{Var}(C_k) &= \sigma_k^2 = \sum_i \sum_j \left( \sigma_r^2[i - k, j]\sigma_s^2[i, j] + \sigma_r^2[i - k, j]m_s^2[i, j] + m_r^2[i - k, j]\sigma_s^2[i, j] \right) \\
\text{Covar}(C_k, C_l) &= \sigma_{k,l}^2 = \sum_i \sum_j \left( \sigma_s^2[i, j]m_r[i - k, j]m_r[i - l, j] + \sigma_r^2[i, j]m_s[i + k, j]m_s[i + l, j] \right)
\end{align*}
\]

Recall that:

\[
\begin{align*}
m_s[i, j] &= f\lambda[i, j] & \sigma_s^2[i, j] &= f^2\lambda[i, j] + \sigma_n^2
\end{align*}
\]
Simplify into meaningful cases

- Combining all those equations is messy and not particularly informative in the general case.

- Let’s examine a few simpler cases to gain insight into the performance of correlation for slope estimation.
  - Assume zero-shifted, identical reference and subimage (closed-loop case).
  - In that case, the means and variances of $C_{-1}, C_1$ are equal.
Image content directly affects variance

- Intuitively, the broader the image’s band of spatial-frequency content, the better the estimation

\[ \sigma_x^2 = \frac{\sigma^2_1 - \sigma^2_{-1,1}}{8(m_0 - m_1)^2} \]

sharpness of autocorrelation
Scene performance depends on frequency content

- Performance varies along x- or y-axis
- Need to find best field of view for expected scene scales
In the case of changing exposure time, the error st. dev. follows as inverse power law in SNR.

\[
\sigma_x(f) = \frac{1}{\sqrt{f}} \left( \tilde{\sigma}_1^2 - \tilde{m}_1 - \tilde{\sigma}_{-1,1}^2 \right)^{1/2} \frac{1}{2\sqrt{2}(\tilde{m}_0 - \tilde{m}_1)}
\]
Background does not bias estimate

- Examination of the mean shift expression shows that a uniform background term cancels out entirely.
- Regardless of scene, no background subtraction necessary.

\[
E[\hat{x}_0] = \Delta x + \frac{0.5(m_1 - m_{-1})}{m_{-1} + m_1 - 2m_0}
\]

All $+b$ terms cancel
Scaling laws - angle of observation

- Generate radiometric models for background levels
- Performance falls off with too much background

\[
\sigma_x(b, f) = \frac{\{N b^2 + 2bf^2(\tilde{m}_0 - \tilde{m}_2 + \tilde{t}f^{-1}) + f^3[\tilde{\sigma}_1^2 - (f - 1)f^{-1}\tilde{m}_1 - \tilde{\sigma}_{-1,1}^2]\}^{1/2}}{2\sqrt{2}f^2(\tilde{m}_0 - \tilde{m}_1)}
\]
Let’s take another look at point sources

- Any questions on the correlation of arbitrary images?
- Back to point sources...
How do we apply correlation?

- Use a fixed reference image
  - e.g. *ideal gaussian spot of expected system FWHM*
- Use of fixed reference significantly reduces noise
- There is some dependence on reference choice
  - *don’t want reference clearly too big or too small*
For large N, correlation has significantly less read noise propagation

- Error variance due to noise does not depend on number of pixels

\[
\text{Var}(\hat{x})_n \propto \frac{\sigma_n^2}{8(m_0 - m_1)^2}
\]

- Procedure
  - Take 100 frames of static-aberration WFS data.
  - Estimate slopes for each subap, each frame
  - Analyze variance of slopes through time to get noise
For large $N$, correlation has significantly less read noise propagation

- Based on spot size from data, predict noise propagator using known formula
- Compare to actual rms noise

<table>
<thead>
<tr>
<th></th>
<th>$\text{Noise prop}$</th>
<th>$\sigma_x$</th>
<th>$\sigma_y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>$1.58 \times 10^{-5}$</td>
<td>$3.8 \times 10^{-3}$</td>
<td>$4.23 \times 10^{-3}$</td>
</tr>
<tr>
<td>Centroider</td>
<td>$212 \times 10^{-5}$</td>
<td>$19.6 \times 10^{-3}$</td>
<td>$20.9 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

LLNL/UC Davis Vision system data from Abdul Awwal
Spot deformities and background bias the centroid estimate

Centroid estimate

After thresholding

Interpolated WFS data

Increasing threshold -> ->

LLNL SSHCL system data from Kai LaFortune
Correlation insensitive to deformities and background

- Used same Gaussian reference for all frames
- Since $N = 12$, also read noise improvement

Same subaperture, six different time steps

Correlation estimate

Interpolated WFS data
The quad-cell method suffers from gain changes with spot size

- Quad-cell is 2 x 2 centroid
- Lost light off edge of pixel also biases estimate
- As spot size changes, so does gain
  - leads to increased residual error if there are non-common-path errors and references

Est is 0.28
Est is 0.14

Figure from Marcos van Dam
Using 4x4 pixels ameliorates the gain problem

- 4x4 centroid has a more regular response with spot size than quad-cell.
- 4x4 correlation is even smoother

Plot data based on algorithm statistical performance and spot models
Testing at Lick of correlation on 4 x 4

- Pixels are ~9 times under-Nyquist sampled for diffraction limited spots
- However, laser spot is very large - it extends beyond quad cell area

- Test white light probe, larger red probe, and laser on the sky, estimate system gain for both algorithms
- Take 4K frames telemetry from system to analyze

*Lick Tests done with Dave Palmer, Elinor Gates and Tony Misch*
Gain high for white probe

- Gain is 1.52 from TT steering mirror to slope estimate
- Both algorithms have same gain in linear region
- Correlation is noisier
Gain lower for red laser probe

- Gain is 0.997 from TT steering mirror to slope estimate
- Both algorithms have same gain
- Correlation is noisier
Both gain and noise results make sense

<table>
<thead>
<tr>
<th>Pixel SNRs</th>
<th>Subap 24</th>
<th>Subap 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.2</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>2.3</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>2.3</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>13.5</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>24.7</td>
<td>18.5</td>
</tr>
<tr>
<td></td>
<td>2.7</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>1.7</td>
</tr>
</tbody>
</table>

- Both probes are small enough that extra pixels have very low signal
  - *using those pixels leads to more noise*

- Gain should go down as spot gets bigger
With LGS, quadcell had low gain

- Moved LGS with uplink mirror
- Used best-centered voltage to reference other trials
- Gain is 1.16
With LGS, correlation had high gain

- Low values are due to truncation in code of output, not algorithm error
- Gain-normalized, noise is equivalent to quadcell
- Gain is 2.0
LGS spot size clearly visible in CCD

(independent color scales)

White probe, dim
Laser probe, dim
LGS on sky

Frame 55
Frame 44
Frame 03

SNR of each pixel shown, 100 frames of CCD data
We have probed spot size curve

Need to calibrate relationships between probes, LGS

How much will spot size vary?
There are other options for algorithms

- Thresholding of background pixels
- ‘Shrinking box’ (Vision AO) - necessary to use fractional pixels
- Optimal estimation strategies
  - see van Dam and Lane (JOSA A 17), Sallberg et al (JOSA A 14), etc.
Analyze the slope estimation algorithm!

- Can do detailed probabilistic analysis of algorithm
- This leads to
  - better understanding of its behavior
  - ability to design and test system to certain performance specifications
- Correlation is preferable to centroiding in certain situations, especially
  - many pixels per subaperture with read noise
  - time-varying background