

**Block-Matching Sub-Pixel Motion Estimation
from Noisy, Under-Sampled Frames – An
Empirical Performance Evaluation**

S. Borman, M. A. Robertson, R. L. Stevenson

Department of Electrical Engineering

University of Notre Dame

Wed. 27 January 1999

1. Introduction

- Why: Multi-frame super-resolution image reconstruction
- Multiple, noisy undersampled frames
→ high-resolution frame
- Reliable sub-pixel motion necessary
- How do motion estimators perform under adverse conditions?
- Want: realistic best case performance

- Attack

- Simulate a realistic imaging system
- Use synthetic scene with known motion
- Estimate motion, analyze errors

2. Imaging System Modeling

- Optical System
 - Diffraction limited, circular aperture
 - f/number 2.8
 - $\lambda = 670\text{nm}$
- CCD Focal Plane Array
 - Square pixels with dimension $X_s = 9\mu\text{m}$
 - Active region is entire pixel
 - Short aperture time
 - Additive Gaussian readout noise

- Sampling Considerations

- Sampling:

- CCD *spacing* determines sampling rate

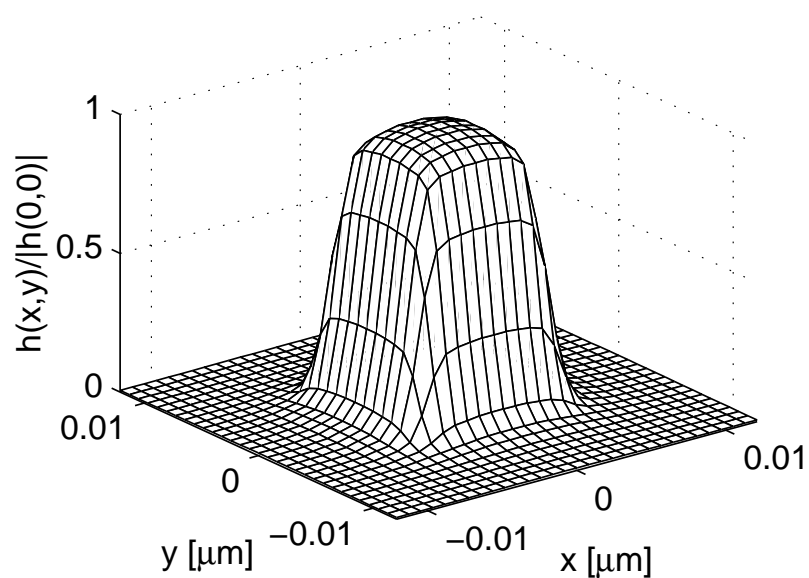
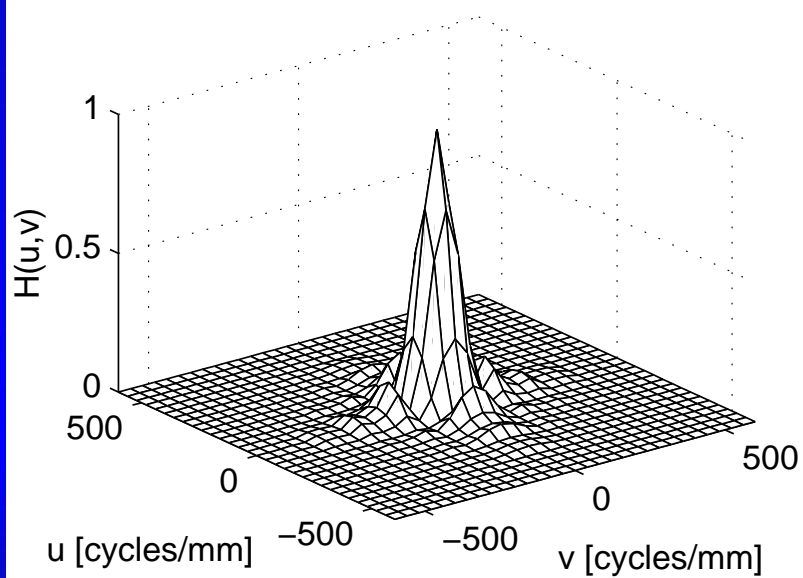
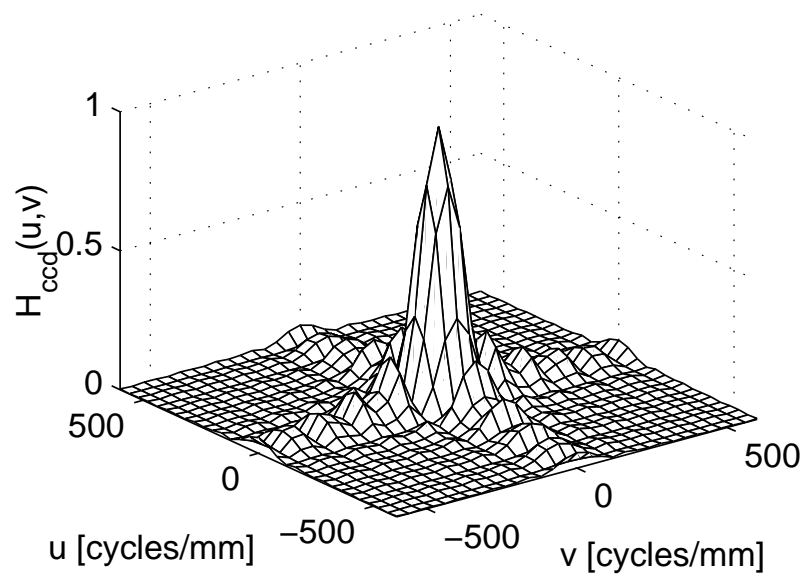
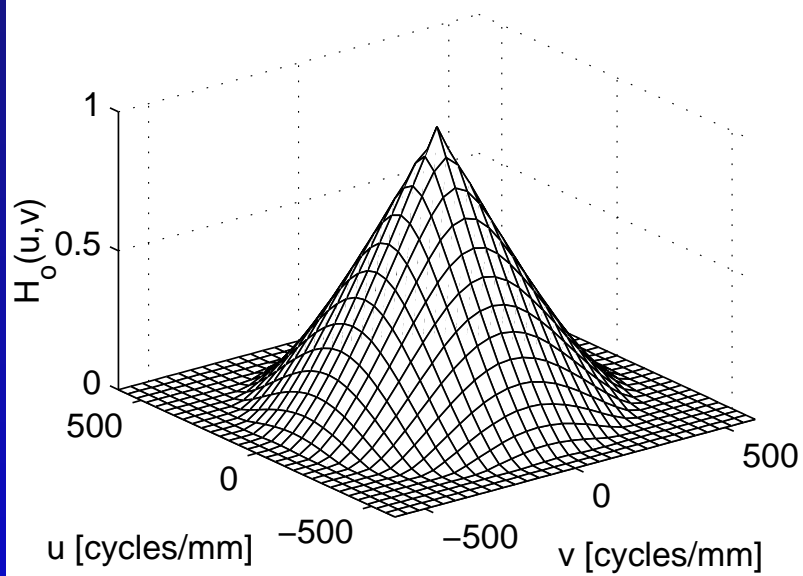
- Not* active region (aperture) of sensor

- Sources of low-pass filtering:

- Optical System response (High f_c)

- CCD spatial integration (sinc, $f_0 = k \frac{1}{X_s}$)

⇒ Undersampled



3. Synthesized Scene Data

- Requirements
 1. Need known scene motion
 2. Seek “realistic best case” performance
 - Consider local translational motion model
 - 1-d “scene”
 - Choose realistic scene (typical in images)
 - Yields very reliable motion estimates

- Continuous-Space Scene Model

- Use step edge $s(x)$, with simple texture $g(x)$

$$f(x, t_1) = s(x) + g(x)$$

$$f(x, t_2) = f(x - X_m, t_1)$$

$$= s(x - X_m) + g(x - X_m)$$

- Translation parameter X_m – to be estimated

- Discrete-Space Approximation

- Sampling of continuous scene model

$$f[k, t_1] = s(k\Delta) + g(k\Delta)$$

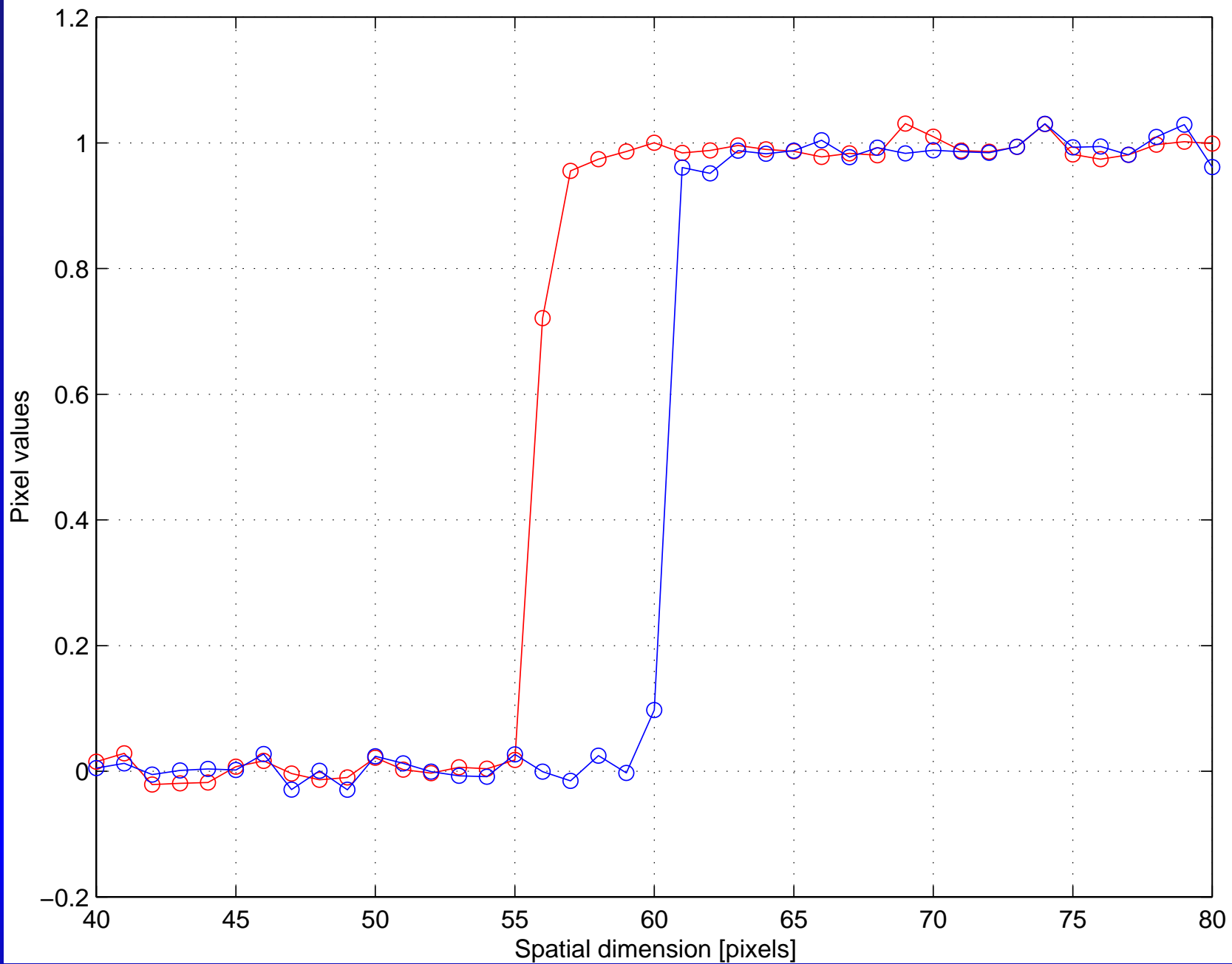
$$f[k, t_2] = f[k - d, t_1]$$

$$= s(k\Delta - X'_m) + g(k\Delta - X'_m)$$

- Δ – sampling rate for discrete-space approximation
- $X'_m \approx X_m$ (Δ -quantized approximation to X_m)
- Translation parameter X'_m – to be estimated

4. Simulated Imaging Process

- Three steps
 1. Low pass filtering (modeling optical system)
 2. Spatial integration and sampling (CCD sensor)
 3. Addition of Gaussian noise (CCD readout noise)
- $f[k, t_1] \longrightarrow \hat{f}[m, t_1]$
- $f[k, t_2] \longrightarrow \hat{f}[m, t_2]$
- In general, $\hat{f}[m, t_2] \neq$ translation of $\hat{f}[m, t_1]$ (texture)



5. Motion Estimation

- Standard block matching motion estimation
⇒ 1-pixel *resolution*
- Interpolate image data $\hat{f}[m, t]$ by a factor of p
⇒ $\hat{f}_I[\frac{i}{p}, t] = \hat{f}[m, t]$ for $i = mp$
- Standard block matching motion estimation on $\hat{f}_I[i, t]$
⇒ $\frac{1}{p}$ -pixel *resolution* estimates
- Perfect: All errors $\pm \frac{1}{2p}$ pixels of true motion
- Resolution vs. Accuracy

- Find the highest motion estimation *accuracy* one can expect to achieve under realistic circumstances
- Choose largest p (highest resolution) such that motion estimates are accurate within some pre-specified bound
- Defn: Accurate to $\frac{1}{p}$ pixels $\Leftrightarrow \alpha$ -% of motion estimation errors fall within $(-\frac{1}{2p}, +\frac{1}{2p})$
- Choose α according to application

- Similarity Measures

$$\text{SAD}(d) = \sum_{i \in \mathcal{B}} \left| \hat{f}_I [i, t_1] - \hat{f}_I [i + d, t_2] \right|$$

$$\text{MSE}(d) = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \left\{ \hat{f}_I [i, t_1] - \hat{f}_I [i + d, t_2] \right\}^2$$

$$\text{NCF}(d) = \frac{\sum_{i \in \mathcal{B}} \hat{f}_I [i, t_1] \hat{f}_I [i + d, t_2]}{\left[\sum_{i \in \mathcal{B}} \hat{f}_I^2 [i, t_1] \right]^{\frac{1}{2}} \left[\sum_{i \in \mathcal{B}} \hat{f}_I^2 [i + d, t_2] \right]^{\frac{1}{2}}}$$

Motion Estimator Test Procedure

1. Choose *a-priori* motion X_m uniformly on $(-wX_s, wX_s)$
 w denotes search window size in pixels

2. Choose X_δ uniformly on $(-\frac{X_s}{2}, +\frac{X_s}{2})$

3. Generate the scene data

$$f[k, t_1] = s(k\Delta + X_\delta) + g(k\Delta + X_\delta)$$

$$f[k, t_2] = s(k\Delta - X_m + X_\delta) + g(k\Delta - X_m + X_\delta)$$

4. Simulate the imaging process $\Rightarrow \hat{f}[m, t_1]$ and $\hat{f}[m, t_2]$

5. Interpolate $\Rightarrow \hat{f}_I[i, t_1]$ and $\hat{f}_I[i, t_2]$

6. Estimate \hat{X}_m from interpolated data

7. Determine motion estimation error $\epsilon = \hat{X}_m - X_m$

5000 Trials for $1 \leq p \leq 20$ and each similarity measure

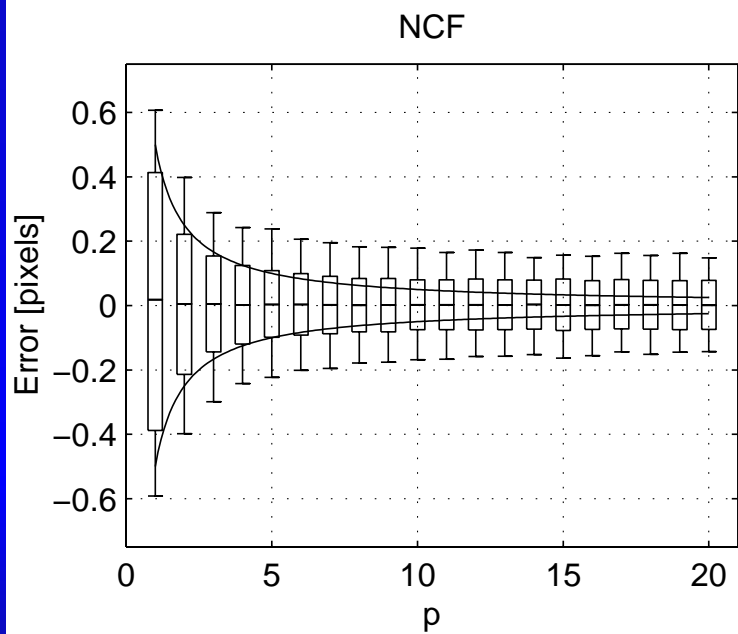
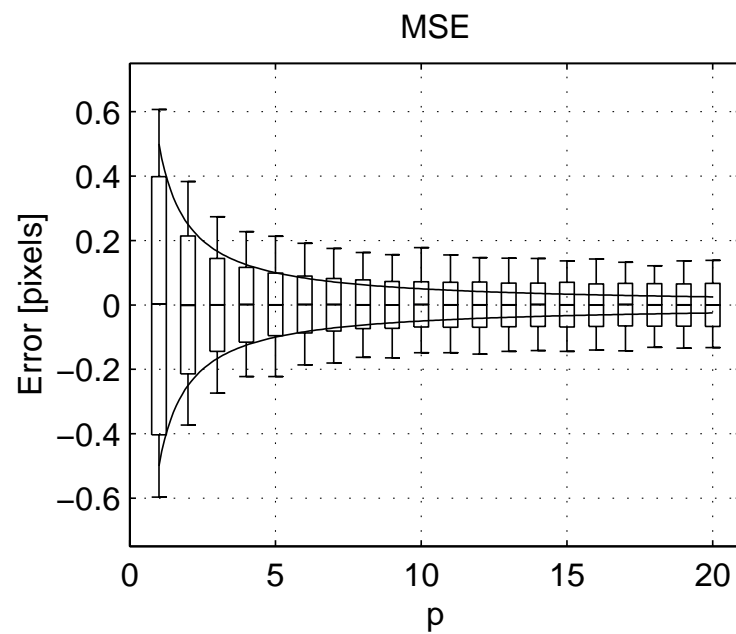
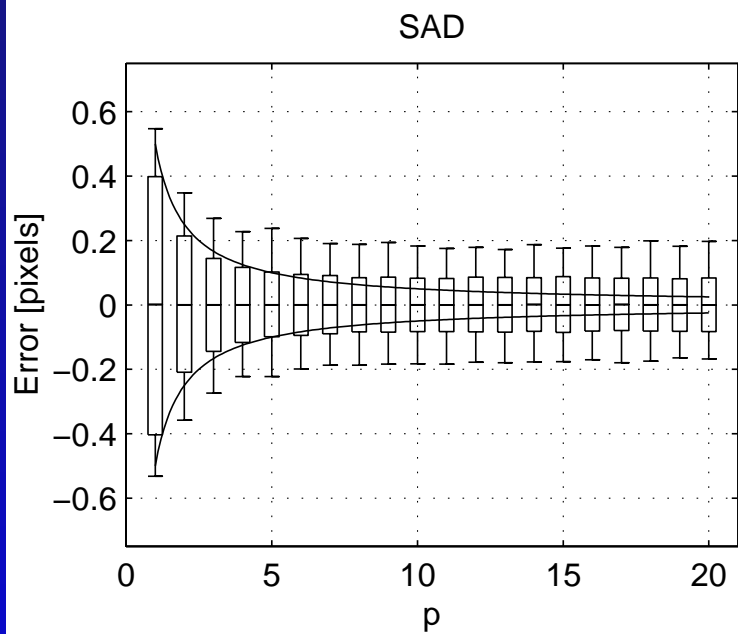
6. Results

1. Fix α

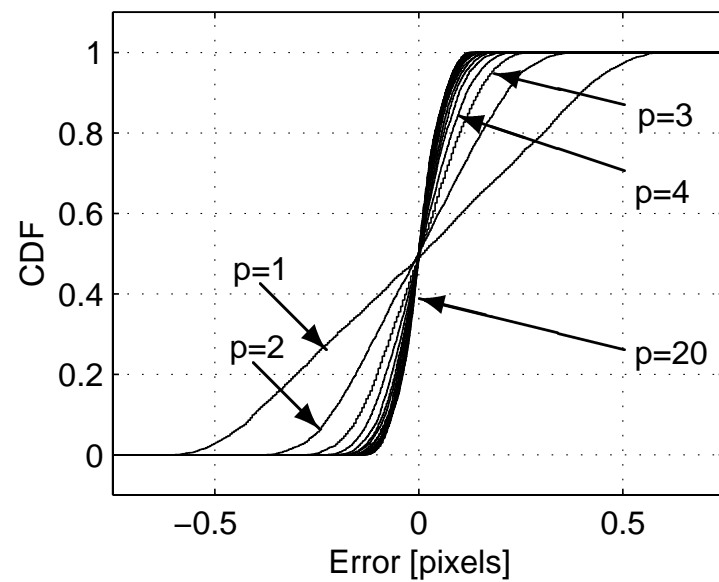
- Compare actual performance with theoretical bounds
- \Rightarrow diminishing returns for increasing p

2. Vary α

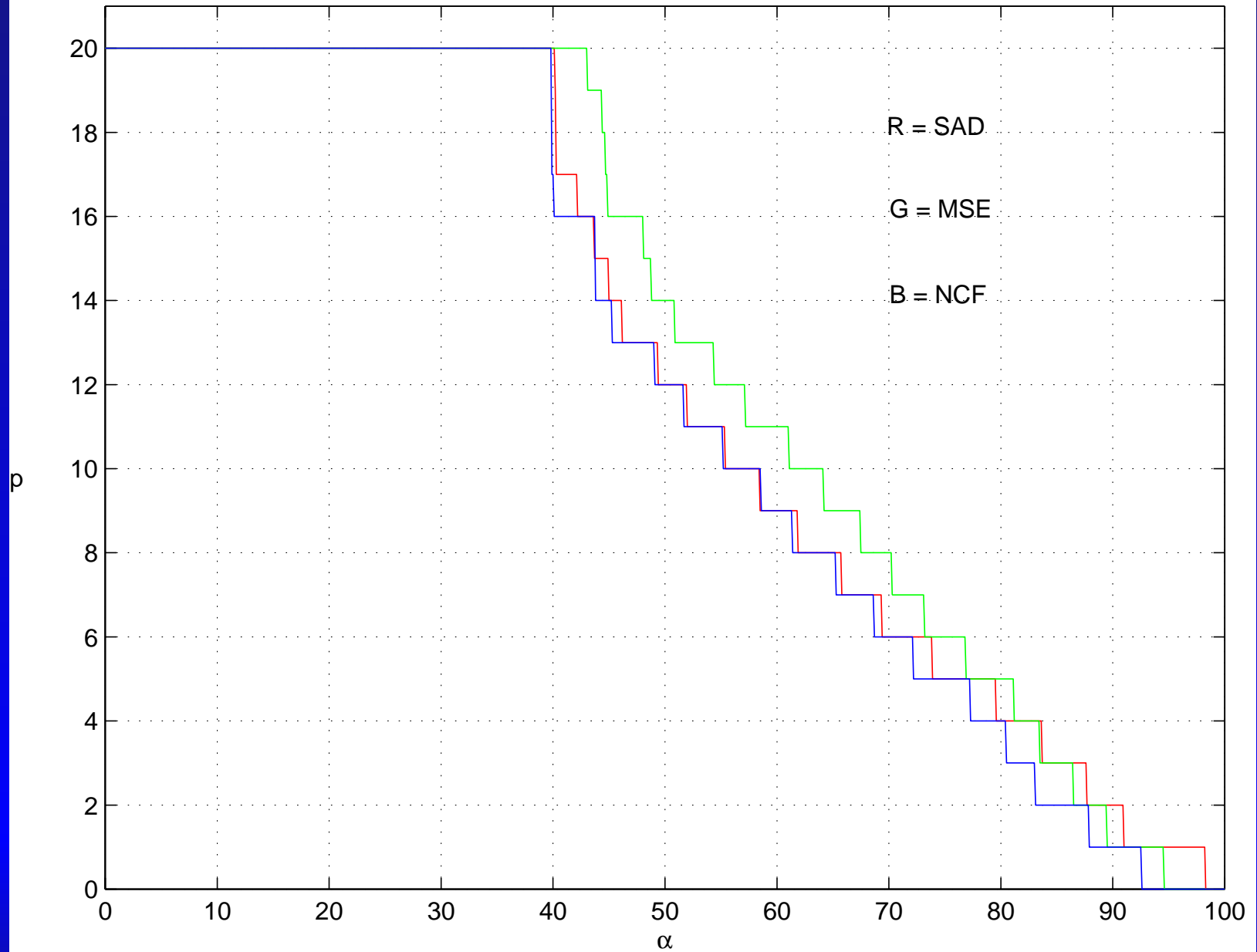
- Find highest p yielding α -% of errors within $\pm \frac{1}{2p}$ pixels of true motion
- \Rightarrow accuracy / resolution trade-off



MSE pixel error cumulative distribution functions



Highest p yielding α -% of motion estimates within theoretical bounds



7. Conclusions

- Seek best case performance of block-matching sub-pixel motion estimators
- Perfect: $\frac{1}{p}$ -pixel motion estimator errors bounded within $\pm \frac{1}{2p}$ pixels of true motion
- Reality: Error approximates $\pm \frac{1}{2p}$ bounds for small p and *levels off* thereafter
- Resolution / Accuracy trade-off
Level of acceptable error determines highest resolution estimator which may be used

Optical System

$$H_o(u, v) = \begin{cases} \frac{2}{\pi} \left\{ \cos^{-1} \left(\frac{\rho}{\rho_c} \right) - \frac{\rho}{\rho_c} \left[1 - \left(\frac{\rho}{\rho_c} \right)^2 \right]^{\frac{1}{2}} \right\}, & \rho < \rho_c \\ 0, & \text{else} \end{cases}$$

$\rho = (u^2 + v^2)^{1/2}$ and u, v spatial frequency variables.

The radial cut-off frequency is given by,

$$\rho_c = \frac{1}{\lambda \cdot f/\text{number}}$$

We choose the wavelength in the visible spectrum as

$\lambda = 670\text{nm}$, resulting in a cut-off at approximately 533 lines per mm.

CCD Focal Plane Array

$$\begin{aligned} h_{CCD}(x, y) &= \frac{1}{X_s^2} \operatorname{rect}\left(\frac{x}{X_s}\right) \operatorname{rect}\left(\frac{y}{X_s}\right) \\ &= \begin{cases} 1/X_s^2, & |x| \leq X_s/2 \text{ and } |y| \leq X_s/2 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

The two-dimensional Fourier transform is given by,

$$\begin{aligned} H_{CCD}(u, v) &= \operatorname{sinc}(X_s u) \operatorname{sinc}(X_s v) \\ &= \frac{\sin(\pi X_s u)}{\pi X_s u} \cdot \frac{\sin(\pi X_s v)}{\pi X_s v} \end{aligned}$$

Sampling Considerations

- Simulated Imaging System

Optical System LPF $f_c = 533$ cycles/mm

CCD spatial integration LPF $f_c = 111$ cycles/mm

CCD sampling rate $f_c = 111$ samples/mm

⇒ Undersampled

Always the case where CCD aperture is anti-alias filter