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

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

Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

ABSTRACT

The performance of block-matching sub-pixel motion estimation algorithms under the adverse conditions of image undersampling and additive noise is studied empirically. This study is motivated by the requirement for reliable sub-pixel accuracy motion estimates for motion compensated observation models used in multi-frame super-resolution image reconstruction. Idealized test functions which include translational scene motion are defined. These functions are sub-sampled and corrupted with additive noise and used as source data for various block-matching sub-pixel motion estimation techniques. Motion estimates computed from this data are compared with the *a-priori* known motion which enables an assessment of the performance of the motion estimators considered.

Keywords: Sub-pixel motion estimation, Super-resolution.

1. INTRODUCTION

Sub-pixel motion estimation techniques are widely used in video coding for transmission and/or compression where half pixel accuracy estimators are routinely used.[1,2] A less well known, but more demanding application is in the emerging field of super-resolution (SR) video and image enhancement where a sequence of *noisy, under-sampled* low-resolution observed frames are used to reconstruct resolution enhanced imagery.[3–6] Accurate *sub-pixel* motion compensation is critical to the success of SR reconstruction algorithms since, (i) motion information provides essential constraints for the solution of the ill-posed SR reconstruction problem, (ii) erroneous motion estimates result in objectionable artifacts after reconstruction and (iii) the degree of resolution enhancement possible is directly determined by the accuracy of the sub-pixel motion information.[7]

Since the sub-pixel motion information required for SR reconstruction must be estimated from the *noisy and under-sampled* observed data, questions concerning the performance of sub-pixel motion estimation techniques under these adverse conditions naturally arise. While it is tempting to apply increasingly higher resolution motion estimation methods in the hope of obtaining improved SR reconstruction, we show that this is impractical.

In order to demonstrate why this is the case, we examine the *"realistic best case"* performance of commonly used block-matching sub-pixel motion estimators. The reason for this is simple — we wish to provided *typical upper bounds* on performance. It is pointless to address worst case performance since this is represented by the case of constant valued regions for which local motion estimation is meaningless. In order to evaluate the "realistic best case" performance, we test motion estimators using synthetic data which are amenable to accurate motion estimation. These data are not *optimal* in the sense that they are the data for which the motion estimators considered yield the *best possible* accuracy, but are chosen so as to represent features of real-world imagery, with the property that, in normal usage, the performance of the motion estimators examined in this paper are typically bounded above by the performance on the synthetic data.

In is important to note that we only address the problem of local motion estimation. We do not consider improvements that are possible using constraints provided by the motion field.

We begin our development by modeling an imaging system consisting of a diffraction limited optical system with a charge coupled device (CCD) focal plane array (FPA) which is subject to additive noise. The synthetic scene data are generated and subjected to known translational motion. The imaging process is then simulated, using the synthetic scenes as input imagery, to generate observation data which closely resemble the output of a typical CCD camera system. Block-matching sub-pixel motion estimation techniques are applied to the simulated camera

Corresponding author: R. L. Stevenson (rls@nd.edu)

imagery resulting in motion estimates which can be compared with the *a-priori* known motion. This approach affords complete knowledge of the true motion and thus enables meaningful statistical evaluation of the performance of the motion estimators.

2. IMAGING SYSTEM MODELING

The imaging system modeled consists of an optical system and a charged coupled device (CCD) focal plane array (FPA) camera operating in optical wavelengths. We discuss each component in turn:

2.1. Optical System

We model a diffraction limited optical system with a circular aperture operating in the visible spectrum with an f/number of 2.8. The modulation transfer function (MTF) for such a system can be shown to be of the form,[8]

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

where $\rho = (u^2+v^2)^{1/2}$ and u, v are spatial frequency variables. The radial cut-off frequency is given by,

$$\rho_c = \frac{1}{\lambda \cdot f/\text{number}}. \tag{2}$$

We choose the wavelength in the visible spectrum as $\lambda = 670$nm, resulting in a cut-off at approximately 533 lines per mm. The optical system MTF is shown in Fig. 2 (a).

2.2. CCD Focal Plane Array

CCD focal plane arrays are ubiquitous in imaging applications and are therefore considered in this work. We model the FPA as a regular array of independent sensing elements, or pixels, which are assumed to be square with edge dimension X_s. Each pixel performs a spatio-temporal integration of the incident illumination over the *active region* of the sensor for the duration of the *aperture time* to yield a measurement. We make several simplifying assumptions concerning the CCD: (i) the active region of each sampling aperture extends to the boundaries delimiting each sensor, (ii) the spatial response over the active region is constant and (iii) the aperture time is of sufficiently short duration so that no blurring due to scene motion occurs. The geometry of the CCD FPA is shown in Fig. 1. In this work, the sampling aperture of each sensor is assumed to have dimensions of 9μm square which is realistic for modern CCD FPA's.

Using the above assumptions, the point spread function (PSF) of a CCD pixel may be modeled as,

$$h_{CCD}(x,y) = \frac{1}{X_s^2}\,\text{rect}\left(\frac{x}{X_s}\right)\text{rect}\left(\frac{y}{X_s}\right) = \begin{cases} 1/X_s^2, & |x| \le X_s/2 \text{ and } |y| \le X_s/2 \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

The two-dimensional Fourier transform is given by,

$$H_{CCD}(u,v) = \text{sinc}(X_s u)\,\text{sinc}(X_s v) = \frac{\sin(\pi X_s u)}{\pi X_s u}\cdot\frac{\sin(\pi X_s v)}{\pi X_s v}. \tag{4}$$

A plot of the magnitude of Eqn. 4 is shown in Fig. 2 (b). This plot also corresponds to the the MTF since the peak at the origin has unit magnitude. Notice that the effect of integration over the active region of the sensor is that of low-pass filtering. Nulls of the spatial frequency response of the sampling aperture occur at integer multiples (excluding zero) of the sampling rate $1/X_s$. The low-pass response of the sampling aperture however, is insufficient as an anti-alias filter since the first null only occurs at the sampling rate. Furthermore, the side-lobes contribute to aliasing.

2.3. Combined Response

The MTF of the combined response of the optical system and a single sampling aperture of the FPA is found by taking the product of the optical system MTF and the sampling aperture MTF. The combined MTF is shown in Fig. 2 (c). The PSF of a CCD pixel, taking into account the response of the optical system, is shown in Fig. 2 (d).

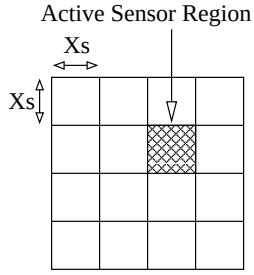


Figure 1. Geometry of CCD Focal Plane Array.

2.4. Sampling Considerations

Next we demonstrate that images captured using typical CCD cameras (as modeled above) fail to meet the Nyquist criterion. We make the realistic assumption that the scene is not band-limited. Two sources of low-pass filtering exist in the imaging system: (i) the diffraction limited optical system with radial frequency cut-off ρ_c in Eqn. 1 and (ii) the low pass sinc filter resulting from the sampling aperture in Eqn. 4.

Consider first the optical system: It is clear from the optical system MTF shown in Fig. 2 (a) that the optical system is a low-pass filter, albeit with a wide pass-band. If the optical system is to be utilized as an anti-aliasing filter, then sampling at the Nyquist rate requires approximately 1066 samples per mm, corresponding to a sensor spacing of approximately 0.94μm. This sampling density is impossible to achieve with current focal plane array technologies. Thus the optical system cannot be relied on to ensure that the image is sufficiently band-limited for critical sampling.

Next consider the CCD FPA: Nyquist rate sampling requires that the image be sampled at twice the highest frequency component present. The CCD FPA sensor spacing determines the sampling rate at $f_s = 1/X_s = 1/9\mu\text{m} \approx 111$ samples per mm. Critical sampling requires that the image be band-limited to $f_{nyquist} = 1/2X_s \approx 55$ cycles per mm. Examining Fig. 2, it is clear that this requirement is not met.

Though the spatial integration performed by the CCD sampling aperture is in effect a low pass filter, it should be noticed that the first zeroes in the frequency response occur (assuming the size of the active area of each sensor is equal to the spacing between sensors) at $f = 1/X_s = 2f_{nyquist}$. Thus the low-pass characteristic of the CCD sampling aperture is insufficient to ensure critical sampling. This argument does not even consider that the CCD aperture response contains multiple side-lobes before the optical system cut-off is reached, nor does it address the fact that for realistic CCD FPA's, there is a "dead zone" between the active regions of adjacent sensors resulting in the first zeroes of the CCD aperture frequency response occurring at higher spatial frequency.

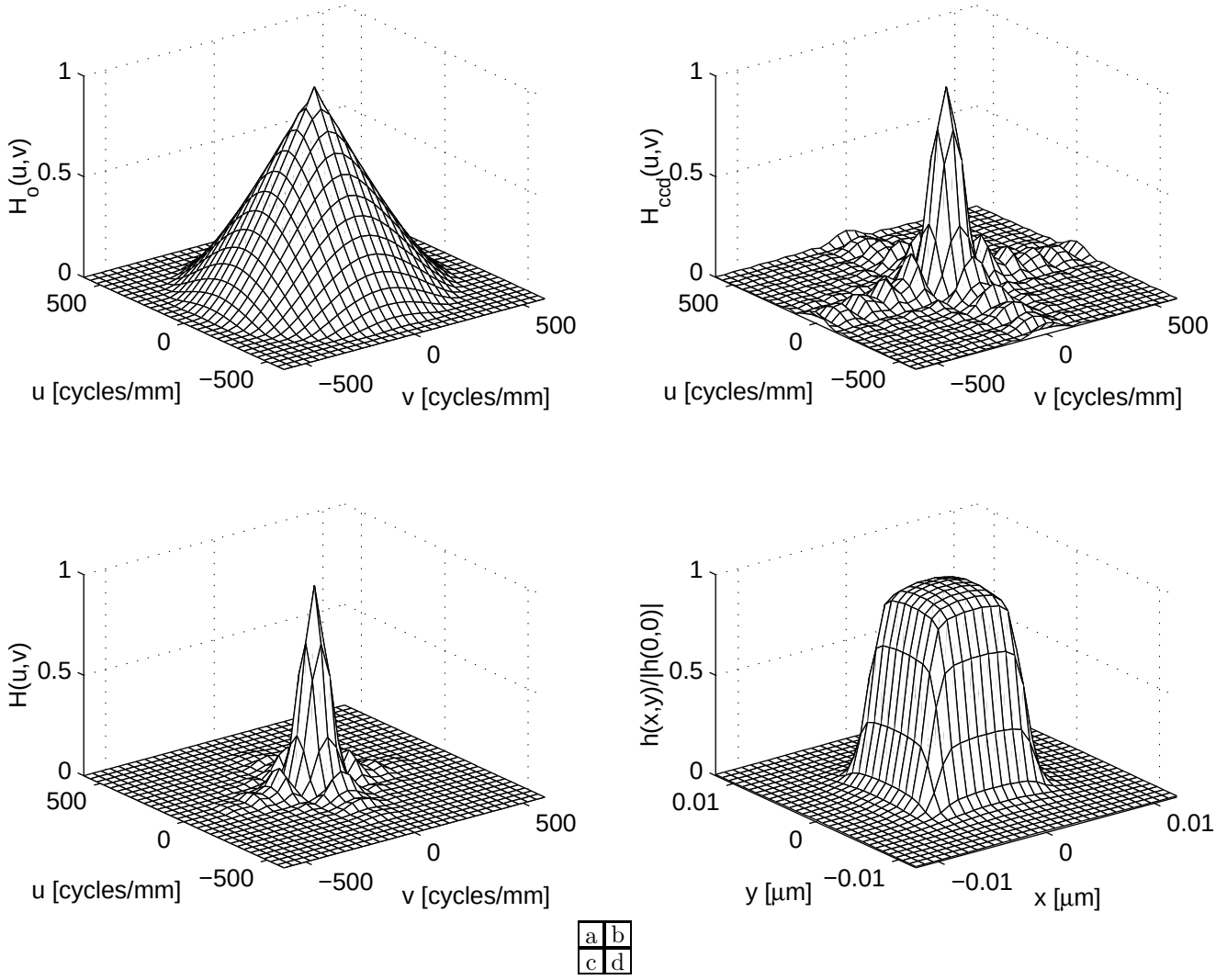
Thus for a typical optical system and CCD configuration, without additional low-pass filtering, aliasing is present. In the imaging system modeled in this paper this undersampling is evident in Fig. 2 part (c), which shows the combined sensor and optical system MTF, in which the MTF is non-zero well beyond the Nyquist rate.

3. SYNTHESIZED SCENE DATA

3.1. Continuous-Space Scene Model

Evaluating the performance of motion estimation algorithms necessitates the use of data for which the motion is known *a-priori*. This precludes the use of uncalibrated real-world data sets where the motion information is not accurately known. In the absence of calibrated sequences, we utilize scenes synthesized using known motion.

Since we are interested in the "realistic best case" performance of block-matching sub-pixel motion estimation techniques, we restrict our attention to local *translational* motion. This is reasonable since (i) estimating translational motion is typically the "easiest" motion to estimate and is most likely to yield the "best case" performance, (ii) while real-world scene motion is obviously not restricted to pure translations, many classes of motion are reasonably approximated by local translational motion, (iii) translational block-motion estimators are quite common and therefore relevant, (iv) the methodology we develop may be applied to other classes of motion.



a	b
c	d

Figure 2. (a) MTF of optical system. (b) MTF of sampling aperture. (c) Combined MTF of optical system and sampling aperture. (d) Normalized PSF of combined optical system and sampling aperture.

Thus far, the background has been developed in terms of the two spatial dimensions present in imaging systems. For the purposes of experimentation, however, we restrict the problem to the *one* dimensional case. This simplification is justified since we consider block-motion estimation methods applied to scenes exhibiting translational motion where two dimensional methods are a straight-forward extension of those in one-dimension. Additionally, the points this paper addresses are most easily demonstrated using a one-dimensional model. Nevertheless, the results are readily applicable to two-dimensions.

To achieve the "realistic best case" we choose a scene which is both realistic and likely to yield the best possible motion estimates. A natural choice which satisfies both requirements is a step edge — commonly found in real-world scenes and a near-optimal function for motion estimation. In order to make the test scene more realistic, texture is also considered. We assume that the step edge $s(x)$ has texture features simulated using an additive white Gaussian noise process $g(x)$, and so has frequency content beyond that passed by the optical system.

For our motion estimation tests, two underlying continuous-space signals are required: The original signal $f(x, t_1)$ at time t_1, and the shifted signal $f(x, t_2)$ at time t_2. In terms of the step edge and the texture model, the continuous-space scene functions are described by,

$$\begin{aligned} f(x,t_1) &= s(x) + g(x) \\ f(x,t_2) &= f(x - X_m, t_1) \\ &= s(x - X_m) + g(x - X_m), \end{aligned} \tag{5}$$

where the translation parameter X_m is to be estimated using the motion estimation techniques.

3.2. Discrete-Space Approximation

For computational purposes, the continuous-space functions $f(x, t_1)$ and $f(x, t_2)$ must be approximated on a discrete space. This is achieved by sampling. Since the highest motion-estimation resolution considered in this work is $\frac{1}{20}$-th pixel (motion estimates quantized to $\frac{1}{20}$-th pixel increments), the sampling period used for the discrete approximation of the continuous-space functions is kept significantly smaller than $\frac{1}{20}X_s$. For the experiments we present here, we utilize a sampling rate of approximately 200 samples per CCD pixel. This rate may at a first glance appear excessively large, however it is necessary in order to obtain a statistically meaningful spread of motion estimation errors. This may be understood with the realization that this sampling rate effectively quantizes the *a-priori* motion to $\frac{1}{200}$-th pixel resolution.

The discrete-space step is first created, and the texture added. This results in the discrete-space approximations to the continuous-space scene functions of Eqn. 5 as,

$$\begin{aligned} f[k,t_1] &= s(k\Delta) + g(k\Delta) \\ f[k,t_2] &= f[k - d, t_1] \\ &= f(k\Delta - d\Delta, t_1) \\ &= f(k\Delta - X'_m, t_1) \\ &= s(k\Delta - X'_m) + g(k\Delta - X'_m), \end{aligned} \tag{6}$$

where Δ is the sampling rate used in the discrete-space approximation, k and d are integers, and $d\Delta = X'_m$ is the Δ-quantized approximation to X_m. For convenience we will refer to X'_m as X_m, and trust that the reader will remain cognizant of this mild abuse of notation.

4. SIMULATED IMAGING PROCESS

The imaging system is simulated in three steps: (i) low-pass filtering to approximate the optical system MTF of Eqn. 1, followed by (ii) numerical integration over the length of each CCD pixel to model the effect of the spatial integration performed by the CCD sensors (the PSF of Eqn. 3) and (iii) addition of Gaussian noise to model CCD readout noise. Using the discrete-space approximation of the synthetic scene as the input to the simulated imaging process results in output which approximates that of a typical CCD camera observing the continuous-space synthetic scene.

The simulated imaging process applied to the discrete-space scenes $f[k, t_1]$ and $f[k, t_2]$ results in the observed discrete-space image data $\hat{f}[m, t_1]$ and $\hat{f}[m, t_2]$, in which each sample represents the output of a single CCD pixel. Samples over the discrete variable m represent a physical distance equal to X_s in continuous-space. Due to the effects

of the optical system MTF, the CCD integration, the added capture noise, and sampling, $\hat{f}[m,t_1]$ and $\hat{f}[m,t_2]$ are in general not equal to the samples of the original continuous-space signals $f(mX_s,t_1)$ and $f(mX_s,t_2)$, respectively.

It is important to realize that due to the spatial integration performed by the CCD pixels, the signals $\hat{f}[m,t_1]$ and $\hat{f}[m,t_2]$ are in general *not* shifted versions of each other. Equality will hold only when the true motion X_m is an integer multiple of the sampling period X_s. This is a consequence of the integration over shifted regions of the texture. After the addition of CCD readout noise independently to each signal however, equality can no longer hold.

A realization of the signals $\hat{f}[m,t_1]$ and $\hat{f}[m,t_2]$ with a shift of 42μm or 4.67 pixels, is shown in Fig. 4 (d). The CCD signal to noise ratio is set at 50dB.

5. MOTION ESTIMATION

With simulated CCD array data derived from a synthetic scene available we are in a position to conduct the motion estimation tests. We begin by reviewing the motion estimators used in this work and move on to discuss the test procedure.

5.1. Motion Estimators Considered

Sub-pixel resolution motion estimation is often achieved using standard block-matching motion estimation techniques performed on *interpolated* image data. Since standard block motion estimation is accurate in theory to 1 pixel, interpolating the image data $\hat{f}[m,t]$ by a factor of p such that $\hat{f}_I[\frac{i}{p},t]=\hat{f}[m,t]$ for $i=mp$ and performing motion estimation on the interpolated data $\hat{f}_I[i,t]$ yields $\frac{1}{p}$-pixel *resolution* motion estimates. This does not, however, mean to say that the motion estimated from the interpolated data are *accurate* to $\frac{1}{p}$-pixel. The raison d'être of this work is to *determine the bounds on the accuracy* that can be achieved using interpolation-based block matching motion estimation.

Block matching motion estimation methods utilize a similarity measure to determine the best match between the reference and displaced image data. We consider three matching criteria[9]: the sum of absolute differences (SAD),

$$\mathrm{SAD}(d)=\sum_{i\in\mathcal{B}}\left|\hat{f}_I\left[i,t_1\right]-\hat{f}_I\left[i+d,t_2\right]\right|, \tag{7}$$

the mean-squared error (MSE),

$$\mathrm{MSE}(d)=\frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}}\left\{\hat{f}_I\left[i,t_1\right]-\hat{f}_I\left[i+d,t_2\right]\right\}^2, \tag{8}$$

and the normalized cross-correlation function (NCF),

$$\mathrm{NCF}(d)=\frac{\sum_{i\in\mathcal{B}}\hat{f}_I\left[i,t_1\right]\hat{f}_I\left[i+d,t_2\right]}{\left[\sum_{i\in\mathcal{B}}\hat{f}_I^2\left[i,t_1\right]\right]^{\frac{1}{2}}\left[\sum_{i\in\mathcal{B}}\hat{f}_I^2\left[i+d,t_2\right]\right]^{\frac{1}{2}}}. \tag{9}$$

The summations are over $\mathcal{B}$, which is a predefined block containing $|\mathcal{B}|$ elements. The motion estimate for each of these criteria is the d which optimizes the matching criteria for $d\in\mathcal{W}$, where $\mathcal{W}$ is some predefined search window of allowable motions. Since $\hat{f}_I[i,t]$ is interpolated by a factor of p from the original CCD data, the actual motion returned by the sub-pixel estimators is $\frac{d}{p}$ pixels, or in continuous-space $\frac{d}{p}X_s$.

For interpolation of the test signals, both linear and cubic spline interpolation techniques are explored.

5.2. Motion Estimator Test Procedure

The goal of motion estimation is to estimate the motion X_m between the signal at times t_1 and t_2 to the highest degree of accuracy possible. A $\frac{1}{p}$-pixel motion estimator yields estimates which are quantized to integer multiples of $\frac{1}{p}$, that is, $\frac{1}{p}$-pixel *resolution*. Additionally, if the $\frac{1}{p}$-pixel motion estimator is *perfect* then the motion estimates will be *accurate* to $X_m\pm\frac{1}{2p}$. In reality, however, we cannot expect the *resolution* and *accuracy* to be equivalent.

The purpose of this research is to find the highest motion estimation *accuracy* one can expect to achieve under realistic circumstances. Since the *resolution* of the motion estimator used provides no assurance as to the *accuracy* of the motion estimates, we choose the largest p (thus yielding the highest possible motion estimation resolution) for which the resulting motion estimates are accurate to within some pre-specified statistical bound.

It is thus reasonable to *define* the following criterion for accuracy: a motion estimator is accurate to $\frac{1}{p}$ pixels if α-% of motion estimation errors fall within the theoretical resolution bounds $(-\frac{1}{2p}, \frac{1}{2p})$. α should be chosen according to the acceptable level of error in a given application.

In order to characterize the statistical properties of the motion estimation errors, 5000 trials for each $1 \leq p \leq 20$ were made. Each trial consists of the following steps:

1. Choose the *a-priori* motion X_m uniformly on $(-wX_s, +wX_s)$ where $\pm w$ denotes the motion estimator search window size in pixels.
2. Choose X_δ uniformly on $\left(-\frac{X_s}{2}, +\frac{X_s}{2}\right)$.
3. Generate the scene data described in Sec. 3 to yield,

$$\begin{aligned} f[k, t_1] &= s(k\Delta + X_\delta) + g(k\Delta + X_\delta) \\ f[k, t_2] &= s(k\Delta - X_m + X_\delta) + g(k\Delta - X_m + X_\delta) \end{aligned}$$

4. Simulate the imaging process of Sec. 4 to determine the observed discrete-space image data $\hat{f}[m, t_1]$ and $\hat{f}[m, t_2]$.
5. Determine the estimate $\hat{X}_m$ by performing block-matching sub-pixel motion estimation using the SAD, MSE and NCF similarity measures on the interpolated data $\hat{f}_I[i, t_1]$ and $\hat{f}_I[i, t_2]$.
6. Determine the motion estimation error $\epsilon = \hat{X}_m - X_m$.

These trials generate 5000 independent values of ϵ for each p and for each similarity measure. This number is sufficiently large to determine reliable statistics. X_δ is introduced to remove any possibility of bias resulting from the spatial location of the step edge in the reference signal $\hat{f}[m, t_1]$. Choosing the translational motion X_m within the range of the motion estimator search window ensures that it is possible for the motion estimators to find the correct translation.

Several variables have potential impact on motion estimator performance: The reciprocal resolution p is the most obvious variable. Indeed, determining the largest p for a given α is a primary goal of this research. In the next section it is shown that the range $1 \leq p \leq 20$ is sufficient to demonstrate a leveling-off behavior in the performance of sub-pixel motion estimators. The next variable of interest is the similarity measure. Utilizing three similarity measures, SAD, MSE, and NCF enables a comparison of the performance of motion estimators using these measures. Another closely related variable is the type of interpolation used for the sub-pixel motion estimation. Linear and cubic spline interpolation techniques are explored. The final variable of interest is the block size $|\mathcal{B}|$ used in the block-matching of Eqns. 7, 8, and 9. The nominal $p = 1$ block size was $|\mathcal{B}| = 7$ pixels (effectively $6p + 1$ pixels after interpolation). The $p = 1$ block size was varied from 7 to 25 pixels in 2 pixel increments to determine its effect.

6. RESULTS

We first compare the empirically-determined accuracy of the sub-pixel motion estimators of increasing resolution with the theoretical resolution limits.

Recall that a perfect $\frac{1}{p}$-pixel resolution motion estimator has errors bounded by $\pm\frac{1}{2p}$. Since the *a-priori* motion is taken as a realization of a uniformly distributed random variable, we expect that for a perfect $\frac{1}{p}$-pixel motion estimator, the error should be distributed uniformly on $(-\frac{1}{2p}, \frac{1}{2p})$.

Results of tests using the three matching criteria SAD, MSE, and NCF are shown in Fig. 3 (a), (b), and (c). Motion estimators of resolution from 1 to $\frac{1}{20}$-th pixel ($1 \leq p \leq 20$) are tested. The abscissas show the reciprocal resolution p of the motion estimators, and the ordinates are the experimentally observed errors in the motion estimates. The boxes bound 80% of the motion estimation errors, with the tails delimiting the remaining errors. The center bar in

the boxes denotes the mean error of the motion estimates. The smooth decaying curves in the figure are plots of the error bounds for a perfect $\frac{1}{p}$-pixel motion estimator, $\pm\frac{1}{2p}$. As described in Sec. 5.2, 5000 randomly chosen scene motions were used for each value of p. Linear interpolation and a block size of 7 pixels was used.

Figure 3 indicates that if an 80% success rate is used as the criterion for satisfactory sub-pixel motion estimation, then the SAD, MSE and NCF measure motion estimators with $\frac{1}{4}$-th to $\frac{1}{5}$-th pixel resolution yield the highest resolution for the given error requirement. The MSE measure estimator performs slightly better than both SAD and NCF. Notice that beyond the highest resolution meeting the 80% success rate there is little improvement in motion estimation accuracy.

Figure 3 (d) shows cumulative distribution functions (CDF) of the estimation errors committed by the MSE estimator with $1 \leq p \leq 20$. Since the *a-priori* scene motion is uniformly distributed, we expect the errors for a *perfect* $\frac{1}{p}$-pixel motion estimator to be uniformly distributed in $(-\frac{1}{2p}, \frac{1}{2p})$; this would correspond to a CDF that

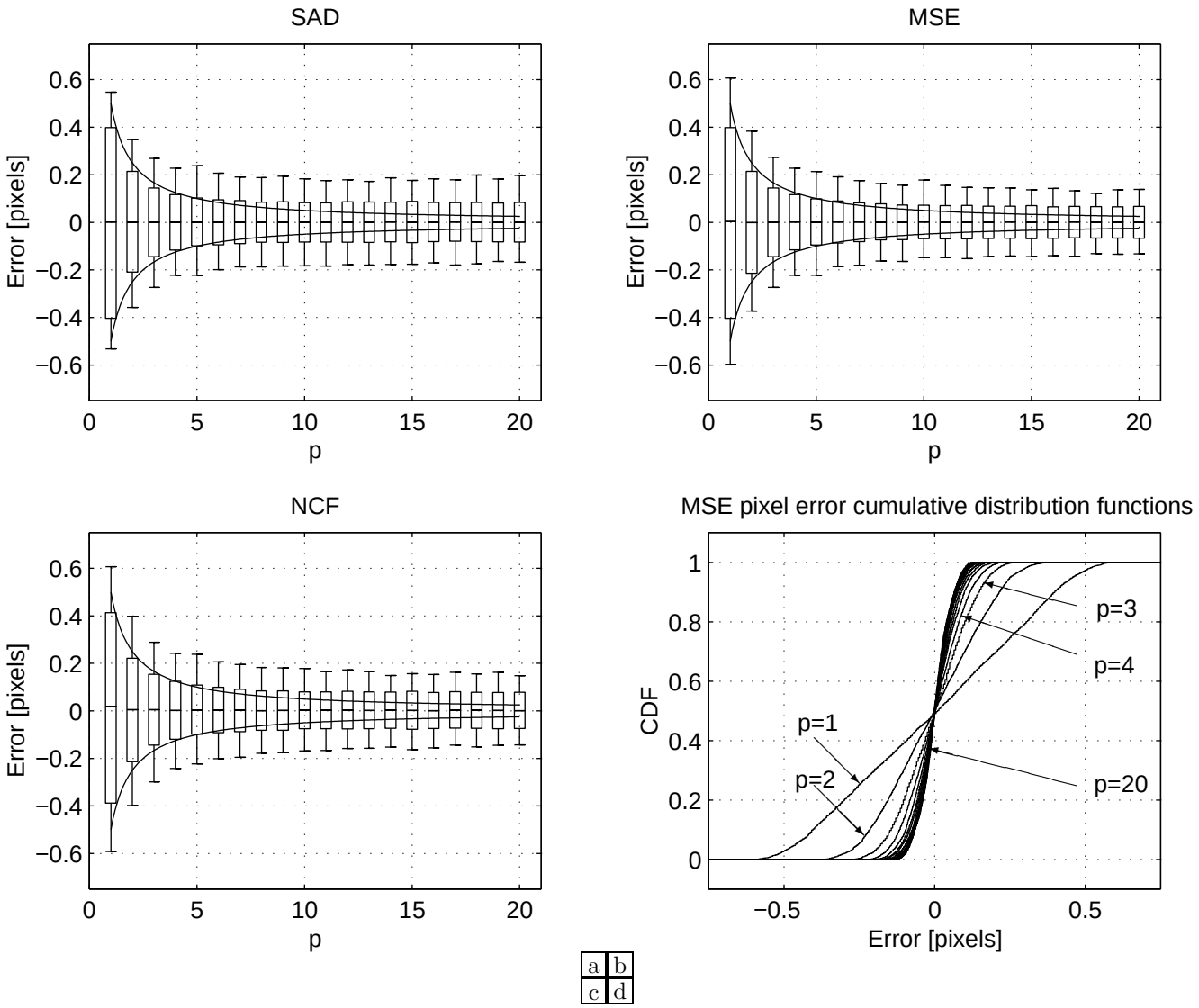


Figure 3. Motion estimation error (in pixels) as a function of increasing estimator resolution. (a) Summed Absolute Differences. (b) Mean Squared Error. (c) Normalized Cross-correlation Function. (d) Cumulative distribution function of errors for MSE motion estimator as a function of interpolation factor.

is linear in $(-\frac{1}{2p}, \frac{1}{2p})$ with slope p. From Fig. 3, it is evident that the CDF is approximately linear, with some rounding occurring at the edges due to errors outside of the $\pm\frac{1}{2p}$ bounds. For values of $p \leq 5$, the slope of the CDF is approximately p. However, for larger values of p, the CDF ceases to exhibit the near-optimal behavior, and approaches a slope (at the 0-error point) of roughly 7. This is indicative of the leveling-off of the the errors evident in parts (a), (b), and (c) of Fig. 3.

A natural issue to address is the performance of the motion estimators for all error percentages α within theoretical bounds. This is shown in Fig. 4 (a), (b) and (c). These graphs are very useful since they provide bounds on the percentage of errors inside (or outside) the theoretical $\pm\frac{1}{2p}$ bounds for each of the $\frac{1}{p}$-pixel resolution motion estimators discussed. These curves provide the highest resolution motion estimator for α-% of errors within the $\pm\frac{1}{2p}$ bounds, that is, the highest p for a given α.

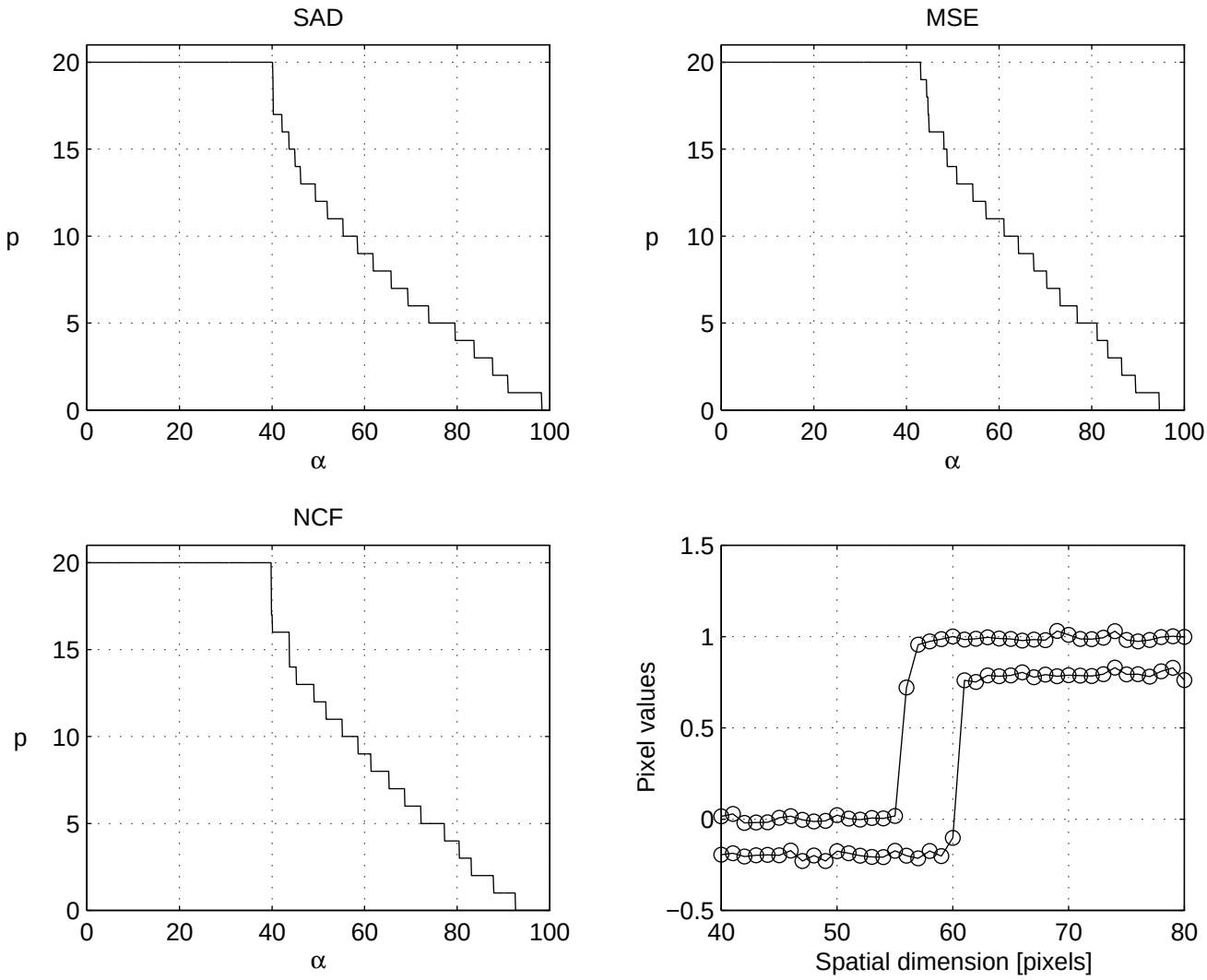


Figure 4. Highest p yielding α-% of motion estimates within theoretical limits. (a) Summed Absolute Differences. (b) Mean Squared Error. (c) Normalized Cross-correlation Function. (d) Typical simulated CCD output for t_1 and t_2. One signal is displayed with a vertical offset for clarity.

The results presented utilize linear interpolation of the CCD data prior to motion estimation. It is interesting to note that there is almost no difference in performance of the motion estimation techniques when using cubic spline interpolation. This does not imply that in general linear interpolation will deliver comparable performance to cubic spline or other more sophisticated interpolation techniques. The similarity in performance observed in these tests is probably a consequence of the choice of test signal; for other more general test functions, using cubic spline interpolation may provide better results.

Also considered is the effect of $\mathcal{B}$, the block over which the similarity measurement is performed. Varying the block size from 7 pixels through 25 pixels in 2 pixel increments yields little difference in motion estimator performance. This should come as no surprise given the step function scene model for which increasing block size will have little effect.

The previous two paragraphs illustrate that the choice of the step function scene model successfully isolates the effects of block size and interpolation method on the test procedure.

7. CONCLUSIONS

This paper began by discussing a model of a typical imaging system, consisting of a diffraction limited optical system with a CCD FPA. A one-dimensional synthetic scene was defined as a step function, with additive noise to represent texture. The test scene was subjected to translational motion. The imaging process was simulated, using the synthetic scenes as input imagery, to generate the discrete-space image data which closely resemble the output of a typical CCD camera system. Block-matching sub-pixel motion estimation using the SAD, MSE, and NCF similarity measures were compared. Both linear and cubic spline interpolation was used. The sub-pixel motion estimation resolution was varied from integer- to $\frac{1}{20}$-pixel.

The differences between the three motion estimation similarity measures were small, but noticeable. The motion estimator using the MSE measure performed best when requiring approximately 90% or less of the motion estimates to be within the theoretical bounds, whereas the motion estimator using the SAD measure was best for above 90%.

The primary result concerns the limits on the accuracy attainable by the sub-pixel motion estimators considered. We showed that a perfect $\frac{1}{p}$-pixel motion estimator exhibits errors bounded within $\pm\frac{1}{2p}$, for $p \geq 1$. For the real-world motion estimators discussed in this paper, the error approximately follows this theoretic bound for *small* p and levels off thereafter. That is, there is a value of the reciprocal resolution p beyond which no additional gains in *accuracy* are made. At this point, it becomes impractical to attempt further increases in p: Increasing p then merely increases the computational complexity of the sub-pixel motion estimation, while the motion estimation accuracy is *not increasing* but rather remains approximately constant. The level of acceptable error in the motion estimates determines the highest resolution motion estimator which may be used. We presented graphs to facilitate this selection.

REFERENCES

1. International Telecommunication Union, *Draft Recommendation H.263+*, Jan. 1998.
2. International Standard ISO/IEC 11172-2, Part 2, *Information Technology—Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s*, 1993.
3. R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing* **5**, pp. 996–1011, June 1996.
4. R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Undersampled Images," *IEEE Transactions on Image Processing* **6**, pp. 1621–1633, Dec. 1997.
5. A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution Video Reconstruction with Arbitrary Sampling Lattices and Nonzero Aperture Time," *IEEE Transactions on Image Processing* **6**, pp. 1064–1076, Aug. 1997.
6. M. Elad and A. Feuer, "Restoration of a Single Superresolution Image from Several Blurred, Noisy, and Undersampled Measured Images," *IEEE Transactions on Image Processing* **6**, pp. 1646–1658, Dec. 1997.
7. R. R. Schultz, L. Meng, and R. L. Stevenson, "Subpixel motion estimation for super-resolution image sequence enhancement," *Journal of Visual Communication and Image Representation, special issue on High-Fidelity Media Processing* **9**, pp. 38–50, Mar. 1998.
8. J. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, 1968.
9. H. Hang and Y. Chou, "Motion Estimation for Image Sequence Compression," in *Handbook of Visual Communications*, H. Hang and J. W. Woods, eds., ch. 5, pp. 147–188, Academic Press, 1995.