

Analysis of Linear Interpolation Schemes for Bi-Level Image Applications

In the office, it is often necessary to scan a picture at a certain resolution and then reproduce it at a different (usually higher) resolution. This conversion can be achieved by interpolating the scanned signal between the sample intervals. This paper discusses a class of linear interpolating methods based on resampling polynomial functions. In addition, we introduce new methods to compare the performance of these interpolating schemes. The signal models used are one-dimensional step and pulse functions. These bi-level models are sufficient to describe many black/white documents. The performance of the linear interpolators is determined by evaluating their accuracy in reconstructing the original bi-level signal. The analysis considers the effects of the coarse scan and fine print intervals as well as the quantization effects. Experiments using the IEEE facsimile chart as input verify the analytical findings. The results show the advantage of using odd-order polynomials, such as the first order and TRW cubic. Also, we discuss the relationship between the interpolating ratio and the number of quantization levels needed to represent the scanned signal.

Introduction

Word processing is becoming widely used to create typed text documents. For more sophisticated documents, such as manuals, diagrams and images are often needed. These images, after being scanned into the system, can go through different phases of editing and formatting by the user. Images may be displayed and printed with devices having different resolutions. In designing a scanning system for this type of application, one must choose the proper resolution for the scanning device. It is most economical to scan the image at the lowest "useful" resolution and then interpolate the image to a higher resolution as needed, provided one does not reduce the apparent image quality when compared with high-resolution scanning. This approach permits the use of a cheaper, low-resolution scanning device with a lower-quality lens system, lower-power light source, fewer electronic buffers, and a lower analog to digital conversion rate. This paper analyzes some performance questions related to two alternative methods of scanning and printing black and white images. The first approach is high-resolution scanning and high-resolution printing (called fine scan/fine print). The other approach is low-resolution scanning of the image, followed by interpolation of the sampled signals to higher-resolution printing (called coarse scan/fine print).

The concept of coarse scan/fine print for black and white images has been investigated by Wong and Schatz [1], who used a nonlinear dot biasing interpolation method. Their approach is simple to implement using an integer coarse scan/fine print ratio, e.g., 2, 3, 4, etc. However, this method is fairly difficult to use for non-integer ratios. Many scanners and printers do not have an integer ratio of resolutions. For document composition, an image is often enlarged or reduced by fractions to fit a fixed space.

Figure 1 represents the scanning and printing system that we analyze here. We assume that the input image is bi-level, e.g., typed characters, line drawings, etc. In the fine scan/fine print (FS/FP) system, the input image is scanned at the high resolution R_1 (spatial interval $T_1 = 1/R_1$), and the scanned signal is then quantized into an appropriate number of levels. If the output device has on-off printing capability only, the stored digitized signal has to be thresholded to either "1" or "0" before printing. The other alternative for reproducing the image is the coarse scan/fine print (CS/FP) system, where the input image is scanned at a lower resolution R_2 (or spatial interval $T_2 = 1/R_2$). The scanned signal is quantized into some appropriate number of levels, and the

© Copyright 1982 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

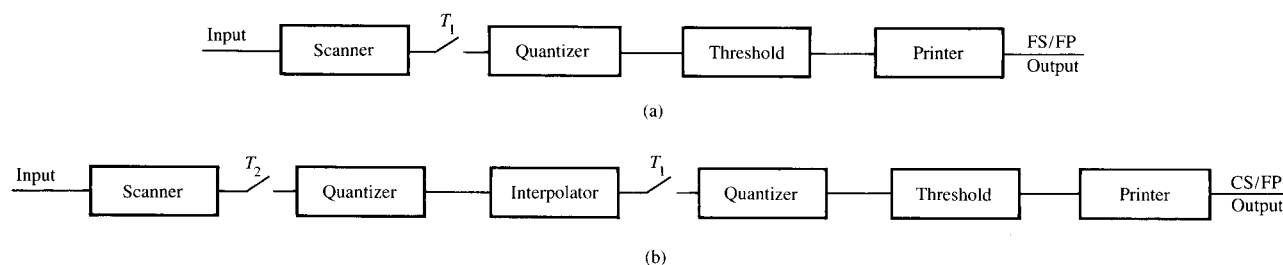


Figure 1 Two configurations of scanning and printing: (a) fine scan/fine print; (b) coarse scan/fine print.

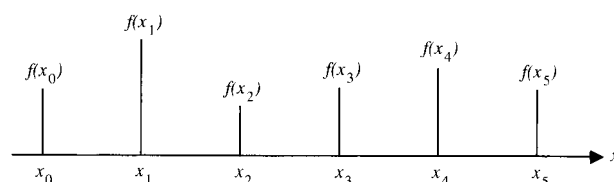


Figure 2 Samples of input signals.

fine-scanned signal is generated by interpolation. The interpolated signal is then thresholded to a two-level signal for the output printer. We are interested in analyzing and comparing the performance of these two methods of reproducing black and white documents. Note that there are two nonlinear elements in both systems, the quantization and the thresholding steps. These nonlinear effects greatly influence the performance of the system.

In the following analysis, we review some linear interpolating schemes and analytical methods that can be used to evaluate the performance of the interpolators. A new performance model is introduced to permit the analysis in the spatial domain. Two input models are used, a step and a pulse input. Detailed analysis is then given for the step input for different system model parameters, *e.g.*, number of quantization levels and coarse scan and fine scan ratios. The analysis is then repeated for the pulse input. Experimental results representing different resolutions of scanning and printing are presented to verify the analysis. Although in this paper we only discuss a one-dimensional signal model, the same analysis can be extended to a two-dimensional model.

Review of linear interpolation functions

The interpolation problem can be stated as follows: Given the values of an unknown function $f(x)$ at a set of equally spaced points $x_0, x_1, x_2, \dots, x_n$, it is required to find an estimate of the function $f(x^*)$ at a point x^* , where $x_0 \leq x^* \leq x_n$ (as shown in Fig. 2).

One way to solve this problem is to pass an n th-order polynomial

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad (1)$$

through the points $(x_0, f(x_0)), \dots, (x_n, f(x_n))$ and to use $P(x)$ as an estimate of $f(x)$. The coefficients $a_0, a_1, \dots, a_n$ can be derived by solving the set of linear simultaneous equations

$$P(x_i) = a_0 + a_1x_i + a_2x_i^2 + \dots + a_nx_i^n = f(x_i) \quad i = 0, 1, \dots, n. \quad (2)$$

Rather than solving a system of linear equations in (2), we may use the Lagrange interpolation polynomial to represent Eq. (1), *i.e.*,

$$P(x) = \sum_{i=0}^n f(x_i)L_i(x), \quad (3)$$

where

$$L_i(x) = \prod_{j=0, j \neq i}^n (x - x_j)/(x_i - x_j) \quad 0 \leq i \leq n. \quad (4)$$

Equation (3) can be formulated as a convolution of an impulse function with the sampled input $f(x_i)$ (see [2]).

As the degree increases, the interpolating polynomial $P(x)$ may not converge to the function $f(x)$ (see [3]). This problem can be solved by using a low-order polynomial to interpolate $f(x)$ on repeated subintervals [3]. Therefore, we consider the interpolating polynomial functions of degrees 0, 1, 2, and 3 only.

• Zero order (nearest neighbor)

Assume that the input signal is the discrete function shown in Fig. 2. A zero-order polynomial is passed through each sampling point. Thus, over the interval $((x_{-1} + x_0)/2, (x_0 + x_1)/2)$, the interpolating polynomial $P(x)$ is determined by solving the equation

$$P(x) = f(x_0) \quad \frac{(x_{-1} + x_0)}{2} < x \leq \frac{(x_0 + x_1)}{2}. \quad (5)$$

The corresponding impulse-convolving function is a one-pixel-width rectangle, as shown in Fig. 3(a).

• First order

In this case we pass a straight line through every two consecutive points of the input signal. Thus, over the interval (x_0, x_1) , the interpolating polynomial is determined by solving the two equations

$$P(x_i) = a_0 + a_1 x_i = f(x_i) \quad i = 0, 1. \quad (6)$$

The corresponding impulse function is a triangle, as shown in Fig. 3(b).

• Second order

In this case a second-order polynomial is passed through three points. Thus, for the interval (x_0, x_2) , the coefficients of the second-order polynomial are determined by the equations

$$P(x_i) = a_0 + a_1 x_i + a_2 x_i^2 = f(x_i) \quad i = 0, 1 \text{ and } 2. \quad (7)$$

It was shown in [2] that the corresponding impulse-convolving function is space-variant with the two repeating impulse responses. In addition to the space-varying property, the second-order interpolator—as all the even-order interpolators—has phase distortion. We discuss other disadvantages of second-order interpolation in later sections.

• Third order

There are many ways of choosing a third-order interpolating polynomial. The “classical polynomial” is chosen so as to pass four points of input signal. Thus, for the interval (x_0, x_3) , we have

$$P(x_i) = a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_i^3 = f(x_i) \quad i = 0, 1, 2, \text{ and } 3. \quad (8)$$

The corresponding impulse response is a space-variant function with three alternating forms. The impulse response has no phase distortion in the central region only (see [2]).

In a “modified classical” approach, we use only the central region of the interpolating function. Thus, there are always two points to the left and two points to the right of the interpolating region. The corresponding impulse response is space-invariant and is described by the equation

$$\begin{aligned} H(x) &= (1-x) \left(1 + \frac{x}{2} - \frac{x^2}{2} \right) & 0 \leq x < 1, \\ &= (1-x)(2-x) \left(\frac{1}{2} - \frac{x}{6} \right) & 1 \leq x < 2, \\ &= 0 & 2 \leq x, \\ H(-x) &= H(x). \end{aligned} \quad (9)$$

The impulse response is shown in Fig. 3(c). Derivation of Eq. (9) is given in the Appendix. The advantages of the modified

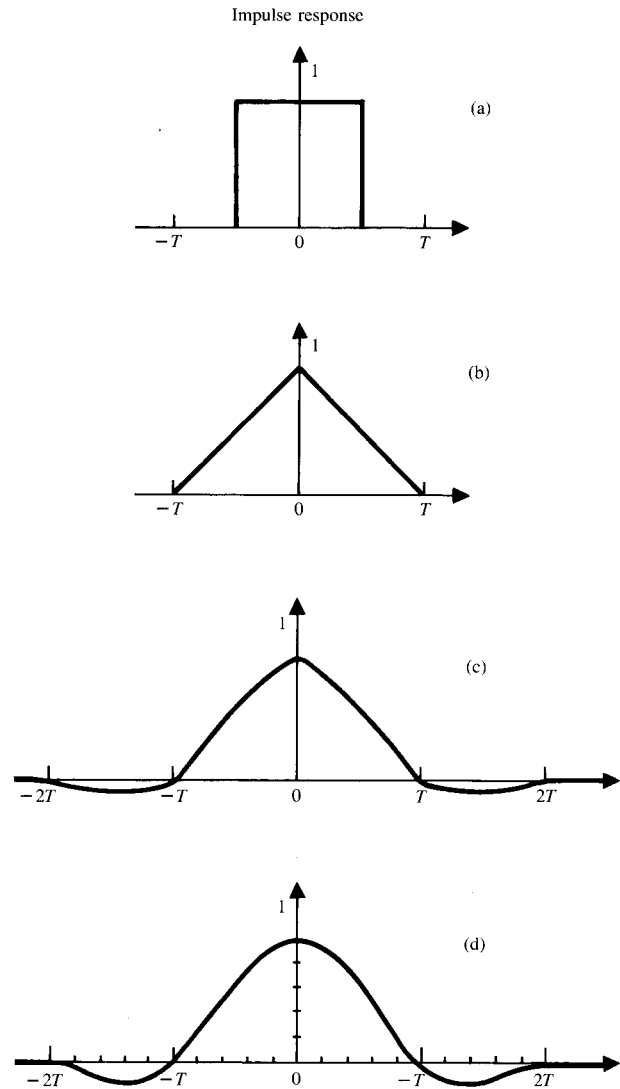


Figure 3 Impulse responses of different interpolators: (a) zero-order interpolation; (b) first-order interpolation; (c) modified third-order interpolation; and (d) TRW cubic interpolation.

cubic polynomial are the space-invariant property, which makes it easier to implement as a linear filter, and the absence of phase distortion in the frequency response (see [2, 4]).

If we are going to use the interpolating function over the central region only, namely (x_1, x_2) , there is no need to pass the interpolating function through the sample points at x_0, x_3 . Instead, these additional degrees of freedom can be used to achieve some other constraints on the interpolating function. A special choice of the cubic polynomial is to approximate the function $\sin x/x$ over the interval $(-2, 2)$ (see [5]). This

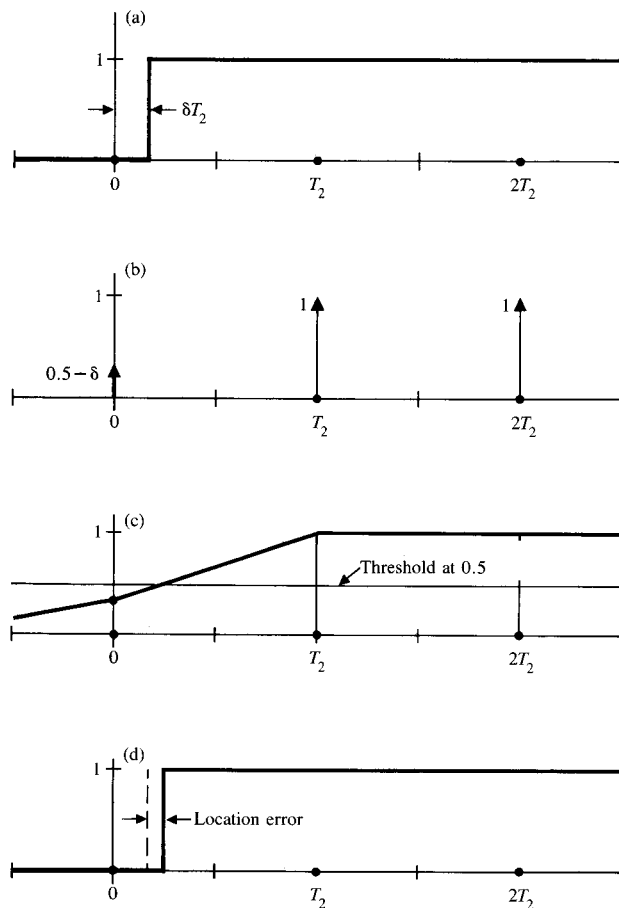


Figure 4 Interpolation of a step function: (a) input step with phase shift δT_2 ; (b) scan at T_2 intervals; (c) first-order interpolation and threshold; and (d) reconstructed step function.

is generally known as the TRW cubic convolution polynomial, and its impulse response is described by

$$\begin{aligned} H(x) &= (1-x)(1+x-x^2) & 0 \leq x < 1, \\ &= (1-x)(2-x)^2 & 1 \leq x < 2, \\ &= 0 & 2 \leq x, \end{aligned}$$

$$H(-x) = H(x). \quad (10)$$

The shape of the impulse response is given in Fig. 3(d). It is shown in the Appendix that the TRW cubic polynomial has the property that the slope of the interpolating function at any sampling point is twice the slope of the line connecting two adjacent points.

Performance evaluation of the interpolating schemes

In the previous section we reviewed some of the polynomial interpolating schemes. We are now going to discuss analytical methods that can be used to evaluate the interpolator's

performance. This evaluation can be done in either the frequency domain or spatial domain.

Examples of the frequency-domain analysis are given in [2, 6]. In [2], Schafer and Rabiner model the interpolating schemes as a digital filter. They assume that the input is made up of ideal impulse samples of a band-limited signal. For this model, ideal interpolation can be achieved using either a low-pass filter or a band stop filter. They calculate the frequency responses of some low-order polynomial interpolations and compare them to the ideal filters. Pratt discusses the problem of reconstructing a continuous signal from the input samples [6] and defines two kinds of reconstruction errors: the resolution error, which describes distortion from attenuation of the central spectra mode; and the interpolation error, which describes the distortion from spurious high-frequency components. These two errors are calculated for a certain band-limited signal model.

Most of the performance evaluation techniques in the spatial domain depend on experimental measurements, such as comparing a reference image sampled at high resolution with the same image interpolated from a coarse-sampled grid. Some interesting measurements are the error histogram, which plots the number of pixels having a given error vs the error magnitude [5] and some functions of this error, such as the mean square error [4].

We previously mentioned that the signals of interest in this paper are binary and that we are interested in analyzing the nonlinear effects of quantization and thresholding. Since binary signals are not band-limited and the nonlinear effects are difficult to analyze in the frequency domain, we do the analysis in the spatial domain. In this approach, the signal models are chosen so that they have practical significance, while being simple enough for the analytical manipulation. Thus, we consider the step and the pulse functions, which can describe almost all binary images of interest. Although we only discuss the one-dimensional signal model, the same analysis can be extended to the two-dimensional model. We now start with the step input model.

• Step input model

Figure 4(a) shows a step input function at a relative position δT_2 from the origin. This input is scanned at intervals T_2 , using an averaging window of T_2 width. The result is shown in Fig. 4(b). The samples are interpolated, in this case using a first-order interpolation for illustrative purposes, and the continuous output signal is shown in Fig. 4(c). To reproduce a binary output, the interpolated signal is thresholded at 0.5 of the intensity, resulting in the output step function shown in Fig. 4(d). The distance between the original and the reconstructed step locations can be used to evaluate the performance of the interpolating function.

In the previous discussion, we neglected two factors, the first being the effect of using a finite number of quantization levels K in representing the scanned signal, and the second being the effect of the fine print sampling interval T_1 . We start with a simplified model and then increase the degree of complexity one step at a time. Thus, we have the following four cases:

1. Neglect the effects of the fine print interval and the quantization error ($T_1 = 0$ and $K = \infty$).
2. Consider the quantization effect (K finite).
3. Consider the effect of the fine print interval ($T_1 \neq 0$).
4. Consider the effects of both the fine print interval and the quantization errors ($T_1 \neq 0$ and K finite).

We start with the first case.

Case 1: Neglecting the effects of fine print interval and quantization error ($T_1 = 0$, $K = \infty$)

The signal model used in this case is illustrated in Fig. 4. We consider some examples for calculating the location error. To simplify the analysis, we normalize the distance measurements by assuming $T_2 = 1$. In the case of first-order interpolation, the reconstructed step location α is at the 0.5 crossing shown in Fig. 5(a). From the triangular ratio we have

$$\alpha = \delta / (\delta + 0.5) \quad (11)$$

$$\text{and the normalized error} = \alpha - \delta. \quad (12)$$

In the modified cubic convolution, the coefficients of the interpolating polynomial are determined by substituting the values of $f(x_{-1})$, $f(x_0)$, $f(x_1)$, $f(x_2)$ in Eq. (8). From Fig. 5(b), we have

$$\begin{aligned} f(-1.0) &= 0 = \alpha_0 - \alpha_1 + \alpha_2 - \alpha_3, \\ f(0.0) &= 0.5 - \delta = \alpha_0, \\ f(1.0) &= 1.0 = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3, \\ f(2.0) &= 1.0 = \alpha_0 + 2\alpha_1 + 4\alpha_2 + 8\alpha_3. \end{aligned} \quad (13)$$

Solving these equations for $\alpha_0, \dots, \alpha_3$ and substituting in $P(x)$, we have

$$P(x) = \left(-\frac{1}{12} - \frac{1}{2}\delta\right)x^3 + \delta x^2 + \left(\frac{7}{12} + \frac{1}{2}\delta\right)x + \frac{1}{2} - \delta. \quad (14)$$

The reconstructed step location will be at

$$P(x) = 0.5. \quad (15)$$

Equation (15) is a cubic polynomial of x and can be solved using Cardan's formula [7] or any simple iterative procedure. It is important to notice that, if we were using the classical cubic interpolation, we would have three forms of $f(x)$, depending on the interval used for interpolation, name-

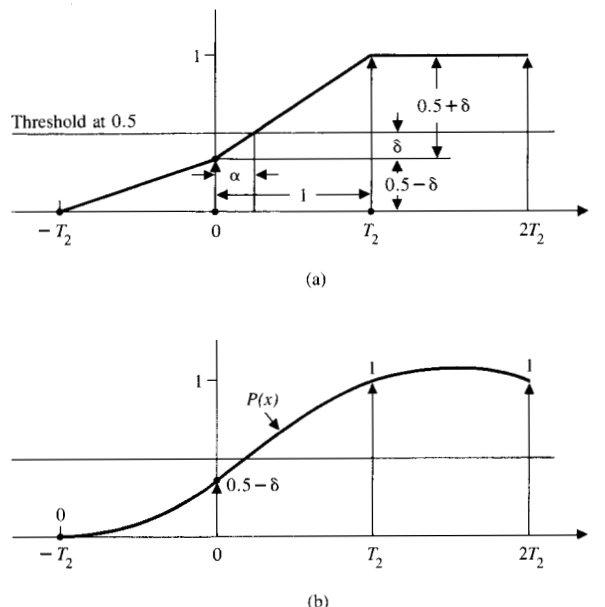


Figure 5 (a) First-order interpolation of a step input; (b) third-order interpolation of a step input.

ly, (x_{-2}, x_1) , (x_{-1}, x_2) , or (x_0, x_3) , with three different values of the location error. Because of this uncertainty, we always use the modified classical form, which we call the "ordinary" cubic polynomial.

The location error as a function of the input step position has been calculated for the zero order, first order, second order, ordinary cubic, and TRW cubic. The results are shown in Fig. 6. From this graph we notice the following:

1. The nearest-neighbor interpolator results in a large location error. The maximum location error is 0.5 of the sampling period, T_2 .
2. The first-order interpolator represents a measurable improvement over the nearest neighbor, the maximum location error being 0.086 of the sampling period, T_2 .
3. For the odd-order polynomials, the error decreases as the polynomial order increases.
4. The TRW cubic has better performance than the ordinary cubic.
5. The even-order polynomials have larger errors than the odd-order polynomials. This results from the lack of symmetry of the even-order polynomials. For a given interpolating point x^* , the number of sampling points to the left and to the right always differ by a factor of one. For this reason we do not consider the even-order polynomials in the rest of this paper.

We now consider the effect of quantization error on the interpolating schemes.

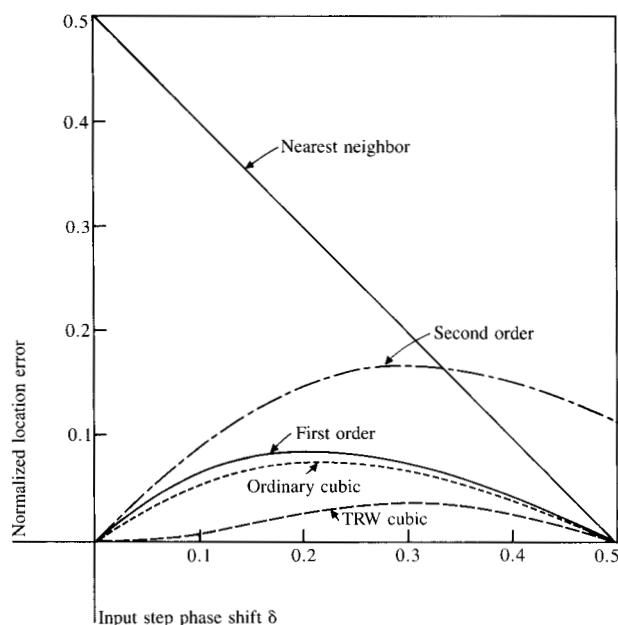


Figure 6 Interpolation errors with step input.

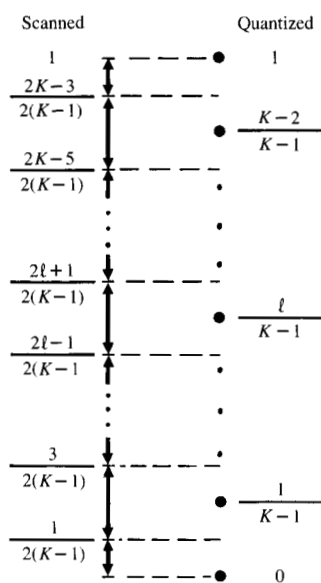


Figure 7 Correspondence between scanned intervals and quantized output for K quantization levels.

Case 2: Considering the effect of quantization error ($T_1 = 0$, K finite)

In this case, the scanned signals are digital with a finite number of quantization levels. Thus, if we have Q bits of quantization with a corresponding $K = 2^Q$ levels, the scanned signal shown on the left of Fig. 7 will be assigned one of these

K levels on the right. In this quantization scheme, we assign levels at 0 and 1 and distribute the other levels uniformly. This scheme minimizes the maximum absolute quantization error. To illustrate the quantization effect, consider the example shown in Fig. 8. We assume two bits of quantization so that we have four possible quantization levels. If the input step is at location δ , such that $0 \leq \delta \leq 1/3$, the scanned signal will correspond to a quantized location $\delta_q = 1/3$. Using first-order interpolation, the reconstructed step location will be at

$$\alpha = \frac{0.5 - \delta_q}{1 - \delta_q}, \quad (16)$$

and the location error is

$$\alpha - \delta = 0.25 - \delta \quad \text{for } 0 \leq \delta \leq \frac{1}{3}. \quad (17)$$

This error is a linear function of δ . A plot of the location error as a function of the step input position is shown in Fig. 9. An important measurement of this graph is the maximum absolute location error over the input step position. We calculate this as a function of the number of bits of quantization for an input step interpolated using the first-order, ordinary cubic, and TRW cubic schemes. The results are plotted in Fig. 10. From this figure, we notice that using one bit of quantization, we have a maximum error of 0.5 the sampling interval. The error decreases monotonically as the number of quantization bits increases, with no significant improvement for $q \geq 5$ bits. Note that TRW cubic convolution has a slightly lower error for the same number of bits of quantization.

Case 3: Considering the effect of fine print interval ($T_1 \neq 0$, and $Q = \infty$)

As an example, assume that the fine print interval T_1 is $1/5$ of the coarse scan interval T_2 , and the input step function is at a position δT_2 from the origin, such that

$$T_1 < \delta T_2 < 2T_1. \quad (18)$$

If the input signal is scanned at the fine rate T_1 , the scanned value will be 0 for $x \leq T_1$ and 1 for $x \geq 2T_1$. The continuous output signal is reconstructed using a rectangular window of width T_1 centered around the sampling point. Thus, the output step will be located at $x = (3/2)T_1$, as shown in Fig. 11(b). The location error is a linear function of the actual step position.

If the input signal is scanned using the coarse interval T_2 interpolated to the fine interval T_1 , the output step position is determined by finding the fine print interval at which the interpolating curve crosses the 0.5 threshold and then reconstructing the step function at the center of this interval. Figures 11(c) and (d) show an example for a CS/FP

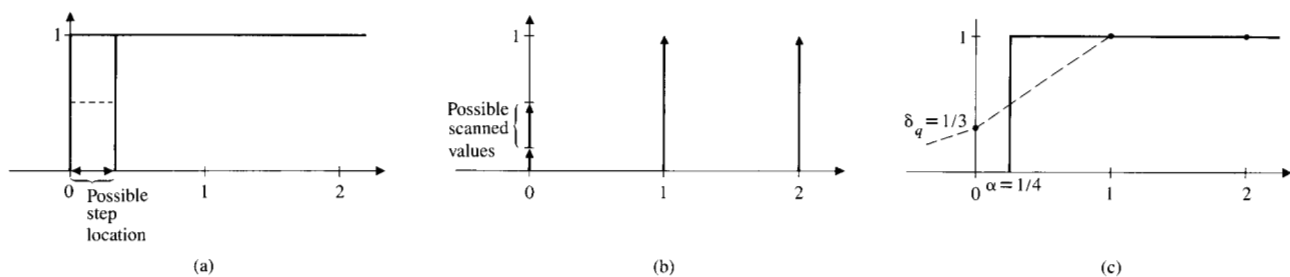


Figure 8 Interpolation of a step function taking into consideration the quantization effect: (a) input step with phase shift $0 \leq \delta \leq 1/3$; (b) scan at unit intervals; and (c) interpolation of quantized signal and threshold in output step.

resolution ratio 1:5 with first-order interpolation. Similar analyses can be applied to the other interpolation schemes. In Fig. 12, we plot the normalized location error as a function of the input step position, for the first order, the ordinary cubic, and the TRW cubic for the CS/FP resolution ratios 1:2, 1:3, 1:4, and 1:5. Notice that the measurements are now normalized with respect to the coarse scan interval T_2 . For comparison, we also plot the results of the FS/FP scheme. These results show that the coarse scan/fine print with different linear interpolation methods and fine scan/fine print have essentially the same form of error curve, which is sawtooth in shape. Furthermore, for most of the input phase shift, all error curves lie on top of each other, with exceptions at few locations. The CS/FP error cannot be reduced below the FS/FP error. It is interesting to note that with a CS/FP ratio of 1:2, the location error is the same whether FS/FP or CS/FP is used with any odd-order interpolator. Some important measurements from the case of a CS/FP ratio of 1:5 are summarized in Table 1.

The above analysis explains why binary images reproduced by using a CS/FP method appear similar to those obtained by a FS/FP method.

Case 4: Considering the effects of fine print interval and quantization error simultaneously

In the analyses of cases 2 and 3, we considered the effects of the quantization error and the fine print resolution separately. We now consider their interaction. We consider the case in which the input scanned signals are represented using 1, 2, 3, and 4 bits of quantization. In this analysis, the scanned signal is quantized to a finite number of levels as in case 2. After interpolation, the step location is determined as in case 3, assuming a CS/FP resolution ratio 1:5. In Fig. 13, we plot the location error as a function of the input step position in the case of first-order interpolation. The results confirm the previous conclusion about the required number of quantization bits. In fact, with $Q = 4$ we have a maximum error less than that for $Q = \infty$. This required number of bits applies to all CS/FP ratios $\leq 1:5$.

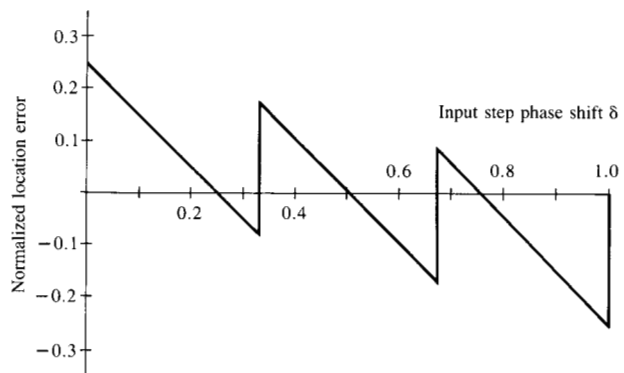


Figure 9 Interpolation error with step input: number of quantization bits, $Q = 2$.

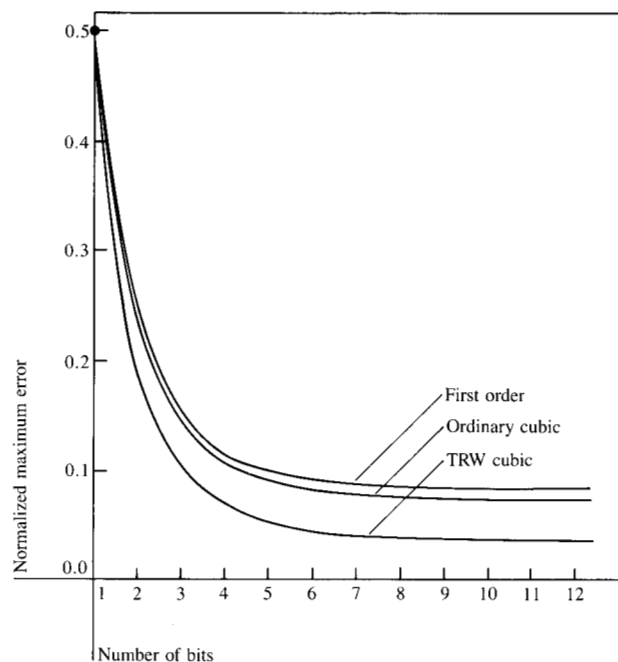


Figure 10 Maximum location error with step input.

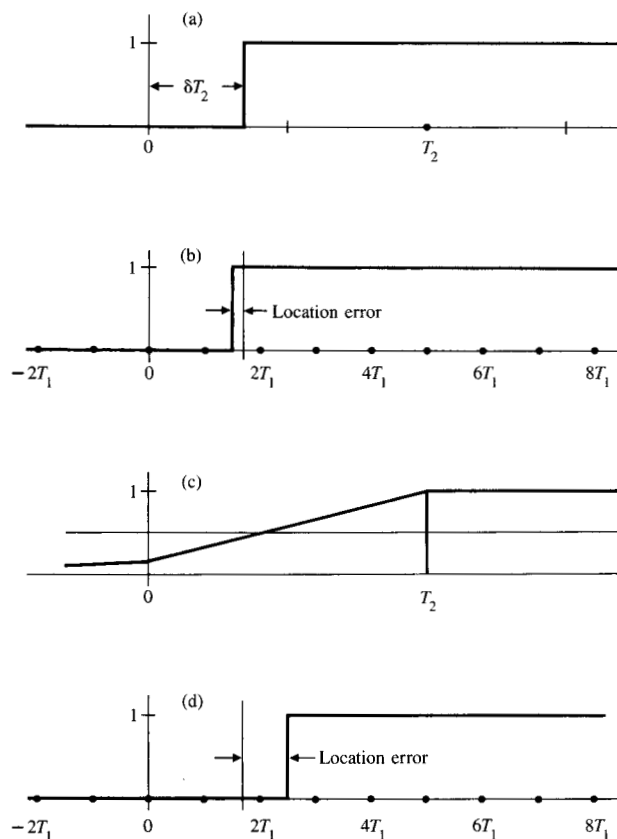


Figure 11 Interpolation of a step function taking into consideration the effect of the fine print interval: (a) input step with phase shift δT_2 ; (b) fine scan at T_1 intervals; (c) interpolation of coarse-scanned signal; and (d) fine print at T_1 intervals.

Table 1 Interpolation errors with step input.

Scheme used	Percentage of region for which the performance of the CS/FP and the FS/FP are the same	Maximum error
FS/FP	100%	$0.1 T_2$
First order	71.4%	$0.175 T_2$
Ordinary cubic	75%	$0.165 T_2$
TRW cubic	89%	$0.13 T_2$

• Pulse signal model

We now discuss the case of an input pulse signal. An example of an input pulse of width W is shown in Fig. 14(a). After scanning, interpolating, and thresholding, the reconstructed pulse signal is as shown in Fig. 14(d). The error measurements of interest for this model are

the positive-going-edge location error $= \epsilon_1 T_2$,

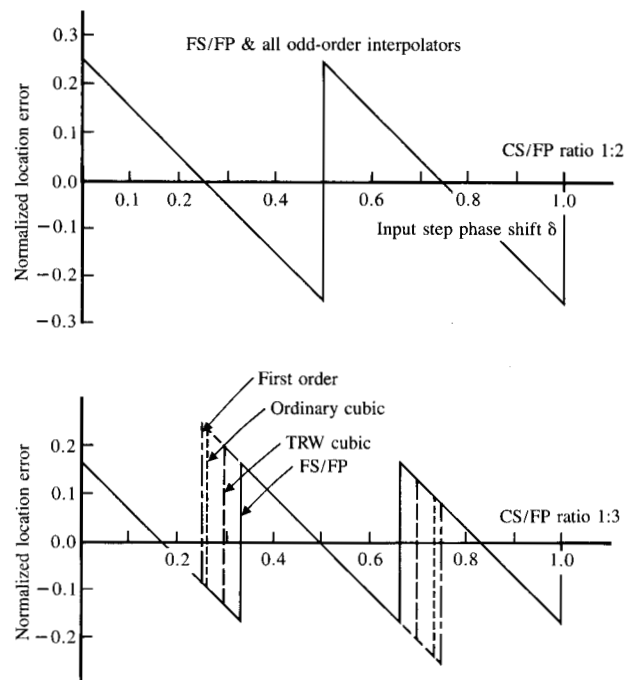


Figure 12 Interpolation errors with step input for different CS/FP ratios.

the negative-going-edge location error $= \epsilon_2 T_2$,

the error in locating the pulse center $= \epsilon_3 T_2 = (\epsilon_1 + \epsilon_2) T_2 / 2$,

the error in estimating the pulse width $= \hat{W} - W$
 $= (\epsilon_2 - \epsilon_1) T_2$.

These error measurements can be derived from the step function analysis if the pulse function has a width

$W > 2T_2$ when we are using a first-order interpolator,

$W > 3T_2$ when we are using a third-order interpolator.

Otherwise, there will be interaction between the positive- and the negative-going edges. In this case, we have to solve the pulse function problem independently. Also, notice that if $W < T_2$, there is a possibility of not detecting the pulse signal at all after thresholding. For the previous reasons, we only consider the case of $T_2 \leq W \leq 2T_2$. In the following analysis, we consider the simple case where the effects of the quantization error and the finite print resolution are neglected. Later we consider the general case.

Case 1: Neglecting the effects of fine print interval and quantization error ($T_1 = 0, K = \infty$)

The signal model used is shown in Fig. 14. In this model, we have two important parameters—pulse width and pulse

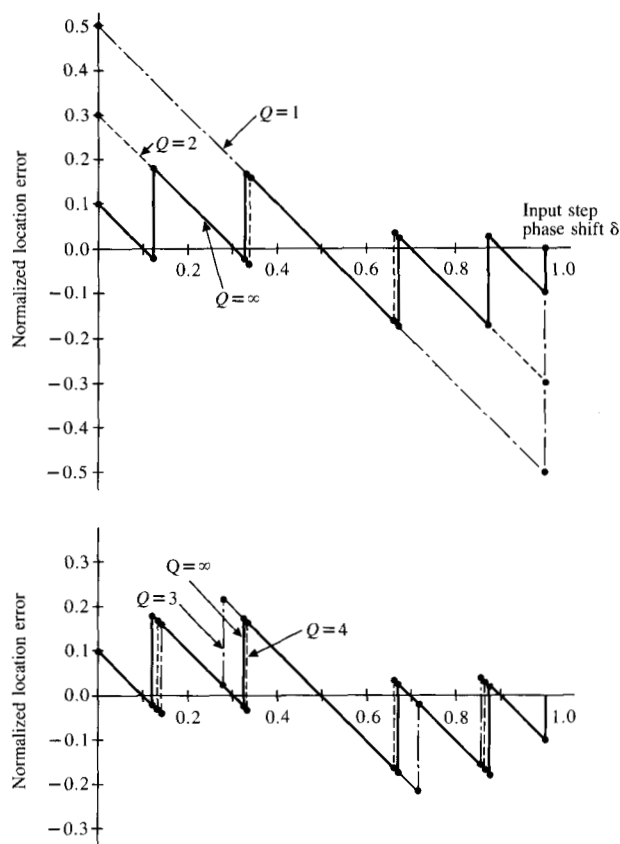


Figure 13 First-order interpolation with different quantization levels and 1:5 CS/FP resolution ratio.

location. For the analysis, we consider input pulses of three widths, $W = 1.2T_2$, $1.5T_2$, and $1.8T_2$. In each case, the scanned signal is interpolated using the first-order, the ordinary cubic, and the TRW cubic methods. Details of the result are described in [8].

For a given input pulse width, the output pulse width is a function of the input pulse position with respect to the sampling intervals. A plot of the maximum and minimum pulse widths for different interpolation methods as functions of the input pulse width is shown in Fig. 15. Notice that the TRW cubic has the smallest error variation, in addition to the fact that most of the TRW error is positive, thus making the pulses appear thicker. This positive error is better than the negative error that may reduce the visibility of some of the thin pulses. Also, we discover that for all the interpolation schemes, the pulse width error increases sharply for $W < 1.2T_2$.

Case 2: Considering the effect of quantization error

We consider the case of an input pulse of width $W = 1.5T_2$. The scanned signal is quantized using 1, 2, and 3 bits of quantization. In each case, the quantized signal is interpo-

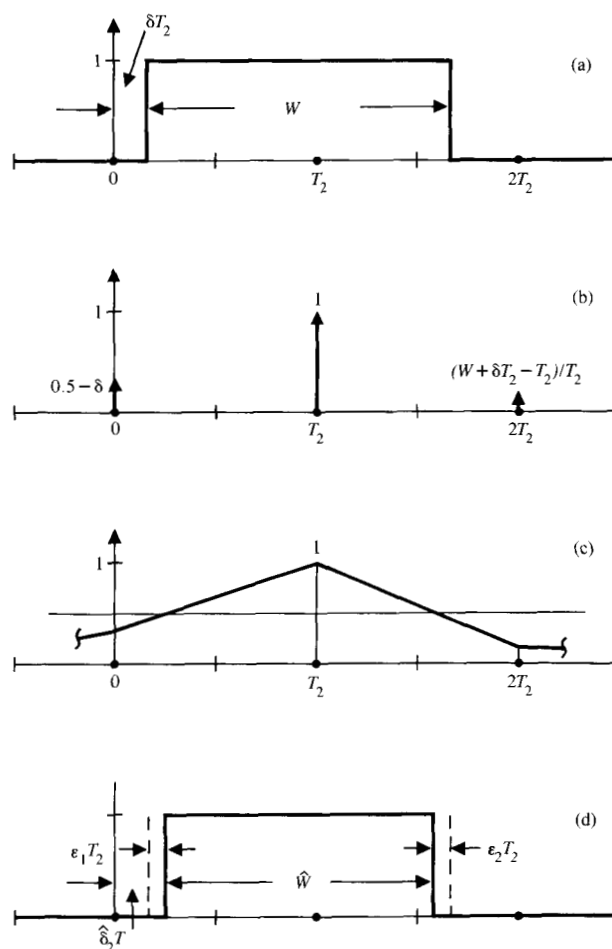


Figure 14 Interpolation of pulse input: (a) input pulse of width W and at position δT_2 ; (b) scan at T_2 intervals; (c) first-order interpolation and threshold; and (d) reconstructed pulse function.

lated using first-order interpolation, and then thresholded. We evaluate the estimated pulse width $\hat{W}$ as the input pulse position varies over the interval $-0.5 \leq \delta \leq 0.5$. Plots of the results are shown in Fig. 16. In the same figure, we plot the pulse error when the quantization error is neglected, namely, $Q = \infty$. From this figure, we notice the large error in the case of $Q = 1$. This error drops sharply for $Q = 2$, and then decreases slowly with the further increase of Q .

Case 3: Considering the effect of fine print interval

As in the previous case, we use as an input a pulse of width $1.5T_2$. We measure the pulse width error when the signal is scanned and printed at a fine interval $T_1 = 1/5T_2$. Also, we measure the pulse width error when the signal is scanned at an interval T_2 and printed at interval T_1 using a first-order interpolation. Plots of the error as a function of the input pulse position are shown in Fig. 17, from which we observe that for most of the input pulse positions, the absolute errors for both the CS/FP and FS/FP schemes are the same.

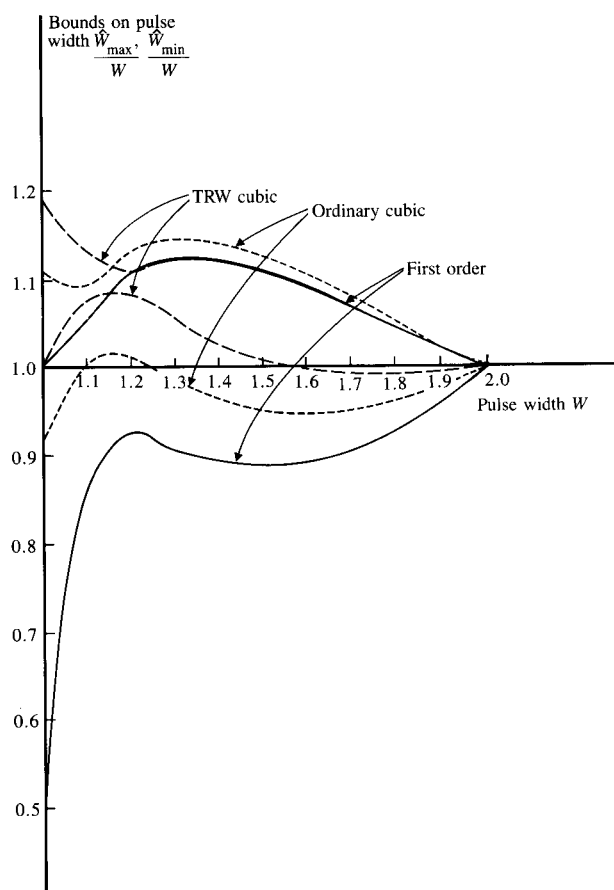


Figure 15 Maximum and minimum pulse width variations.

Experimental results

In the previous sections we discussed different analytical measurements that can be used to evaluate the polynomial interpolating schemes. To verify the analytical findings, we conduct the following two experiments using the IEEE facsimile test chart. In the first experiment we scan the parts corresponding to the "text," "scale," and "ray," at a rate of 1248 pixels/inch, as shown in Fig. 18. The scanned pictures are reduced by a factor of five using a simple averaging procedure. The results are then interpolated five times using the nearest neighbor, the first order, the ordinary cubic, and the TRW cubic. For each test picture, we compare the interpolated output with the original picture. As an error measurement, we count the number of originally white pixels that have been changed to black and the number of black pixels that have been changed to white. The percentage errors are also calculated. The results are given in Tables 2, 3, and 4. In each case, the number of originally white pixels ($sumw$) and the number of originally black pixels ($sumb$) are given in the table.

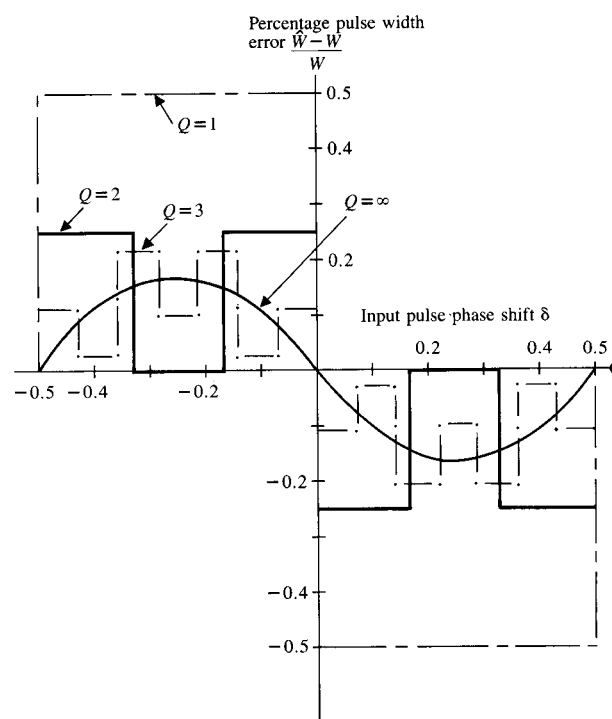


Figure 16 Quantization effects on pulse width variations using first-order interpolation (pulse width $W = 1.5$).

Note that the nearest neighbor results in errors that are at least double those of the other interpolating schemes. The differences between the other schemes are small, with the TRW cubic resulting in the best performance, followed by the ordinary cubic and the first order. These results agree very well with the previous analysis.

In the second experiment, we scan parts of the IEEE chart at a coarse scanning rate of 249 pixels/inch. The scanned signal is then interpolated five times and compared with the fine-scanned signal. Examples of the results are shown in Figs. 19 and 20. In Fig. 19, we show the two-point and six-point letters of the IEEE chart in the cases of FS/FP and CS/FP using the nearest neighbor, the first order, the ordinary cubic, and the TRW cubic. These results show that the nearest neighbor is inferior to the other interpolating schemes. We also observe that there is a limit on the coarse scan rate beyond which it is impossible to regain the original signal. This is clearly shown in the case of the two-point letters, where the stroke width is 0.6 of the coarse scan interval. In this case, none of the interpolated outputs is legible while the FS/FP output is still sharp and clear. Also of interest is the effect of changing the number of quantization levels of the scanned signal on the interpolated output. In this experiment, we scan parts of the IEEE chart at 249

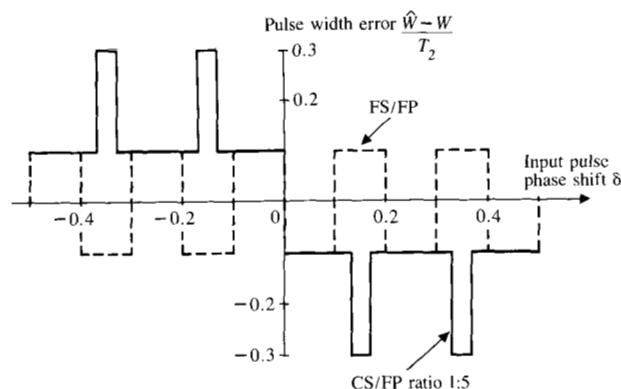


Figure 17 Effect of CS/FP ratio on pulse width error using first-order interpolation.

Table 2 Picture "text": $sumw = 3\,730\,625$; $sumb = 453\,439$.

Operator	Error			Percentage error		
	White	Black	Total	White	Black	Total
NN	52 154	33 578	85 732	1.40	7.41	2.05
B1	25 981	4 450	30 431	0.70	0.98	0.73
ORD	25 349	2 577	27 926	0.68	0.57	0.67
TRW	24 616	1 573	26 189	0.66	0.35	0.63

pixels/inch, while the output is represented using 1, 2, 4, 6, and 8 bits of quantization. These outputs are then interpolated five times and compared with the FS/FP at 1248 pixels/inch. Results for the six-point text are shown in Fig. 20. There is no noticeable difference between the 8-bit, the 6-bit, and the 4-bit pictures. There are some differences between the 4-bit and the 2-bit pictures, while the 1-bit picture is obviously inferior to all the other cases. These observations agree with the analytical results obtained in previous sections.

Conclusions

This paper has presented both analytical and experimental results in evaluating the performance of different linear interpolation schemes for bi-level images. Although we only use a one-dimensional step and pulse input model for our analysis, we believe that our results are directly applicable to binary images, which consist of a combination of step and pulse inputs.

The odd-order interpolators have better performance than the even-order interpolators. The nearest-neighbor method

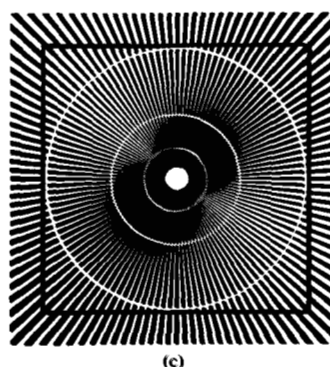
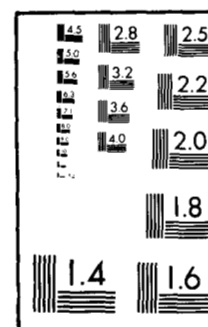
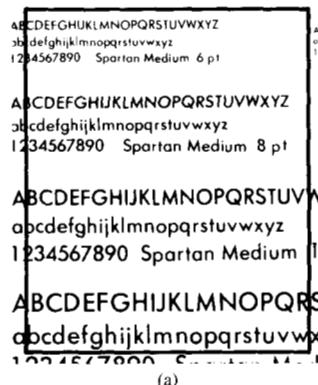


Figure 18 Selected portions of the IEEE Facsimile Test Chart: (a) picture "text," (b) picture "scale," (c) picture "ray."

Table 3 Picture "scale": $sumw = 3\,452\,401$; $sumb = 731\,663$.

Operator	Error			Percentage error		
	White	Black	Total	White	Black	Total
NN	70 848	43 217	114 065	2.05	5.91	2.73
B1	46 525	9 590	56 115	1.35	1.31	1.34
ORD	42 474	7 529	50 003	1.23	1.03	1.20
TRW	40 182	6 404	46 586	1.16	0.88	1.11

Table 4 Picture "ray": $sumw = 2\,224\,873$; $sumb = 1\,959\,191$.

Operator	Error			Percentage error		
	White	Black	Total	White	Black	Total
NN	197 884	116 731	314 615	8.89	5.95	7.52
B1	129 282	7 575	136 857	5.81	0.39	3.27
ORD	119 119	7 126	126 245	5.35	0.36	3.02
TRW	116 770	8 024	119 701	5.02	0.41	2.86

S249/P249 (5X) ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890 Spartan Medium 2 pt

Bi-linear ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890 Spartan Medium 2 pt

TRW cubic ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890 Spartan Medium 2 pt

Ordinary cubic ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890 Spartan Medium 2 pt

S1248/P1248 ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890 Spartan Medium 2 pt
(a)

S249/P249 (5X) nopqrstuvwxyz
Spartan Medium 6 pt

Bi-linear nopqrstuvwxyz
Spartan Medium 6 pt

TRW cubic nopqrstuvwxyz
Spartan Medium 6 pt

Ordinary cubic nopqrstuvwxyz
Spartan Medium 6 pt

S1248/P1248 nopqrstuvwxyz
Spartan Medium 6 pt
(b)

Figure 19 (a) Reproduction of IEEE chart two-point letters; (b) reproduction of chart six-point letters.

(zero order) is simple to implement, because it does not involve arithmetic operations, but its performance is inferior to other odd-order interpolators. Among the odd-order interpolators tested, the TRW cubic has the best performance, followed by the ordinary cubic and then the first order. Since the differences in performance of these three interpolators are not significantly large, we recommend using the bi-linear interpolation method for image resolution translation from low to high resolution, because it requires a smaller number of computation steps than the TRW cubic and the ordinary

cubic interpolators. The bi-linear interpolator needs four points in an image to calculate a new sampling point, whereas the cubic interpolators need sixteen points.

In processing binary images for interpolation, we have shown that quantization of 4 to 5 bits is necessary to preserve very reasonable edge location accuracy.

Appendix

In this appendix, we prove that the modified cubic and the TRW cubic interpolators can be modeled as a convolution with an impulse response of the form shown in Figs. 3(c) and 3(d), respectively. In general, this impulse response is defined as

$$\begin{aligned} H(x) &= h_1(x) & 0 \leq x < 1, \\ &= h_2(x) & 1 \leq x < 2, \\ &= 0 & 2 \leq x, \end{aligned}$$

$$H(-x) = H(x), \quad (19)$$

where both h_1 and h_2 are third-order polynomials. If the input samples are as shown in Fig. 21, the interpolated output over the central region, (0, 1), is the sum of the impulse responses from $f(-1)$, $f(0)$, $f(1)$, and $f(2)$. Thus, the interpolating polynomial is

$$\begin{aligned} P(x) &= f(-1)h_2(x+1) + f(0)h_1(x) \\ &\quad + f(1)h_1(1-x) + f(2)h_2(2-x). \end{aligned} \quad (20)$$

Since the interpolating function has to pass through the sampling points, $h_1(x)$ and $h_2(x)$ satisfy the following conditions:

$$h_1(0) = 1, \quad (21)$$

$$h_1(1) = 0,$$

$$h_2(1) = 0,$$

$$h_2(2) = 0. \quad (22)$$

Thus, the polynomials h_1 and h_2 are of the form

$$h_1(x) = (1-x)(1+Ax+Bx^2), \quad (23)$$

$$h_2(x) = (1-x)(2-x)(C+Dx). \quad (24)$$

The parameters A , B , C , and D are determined by the interpolating scheme used. We now consider two special cases.

In the modified cubic interpolator, the function $P(x)$ and its extensions over the regions $(-1, 0)$ and $(1, 2)$ is a third-order polynomial of the form

$$P(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 \quad (25)$$

that should pass through all the sampling points. Thus,

$$\begin{aligned}
P(-1) &= f(-1) = \alpha_0 - \alpha_1 + \alpha_2 - \alpha_3, \\
P(0) &= f(0) = \alpha_0, \\
P(1) &= f(1) = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3, \\
P(2) &= f(2) = \alpha_0 + 2\alpha_1 + 4\alpha_2 + 8\alpha_3.
\end{aligned} \quad (26)$$

Solving for $\alpha_0, \alpha_1, \alpha_2$, and α_3 , and substituting in Eq. (25), we have

$$\begin{aligned}
P(x) &= f(0) \\
&+ \left[-\left(\frac{1}{3}\right)f(-1) - \left(\frac{1}{2}\right)f(0) + f(1) - \left(\frac{1}{6}\right)f(2) \right]x \\
&+ \left[\left(\frac{1}{2}\right)f(-1) - f(0) + \left(\frac{1}{2}\right)f(1) \right]x^2 \\
&+ \left[-\left(\frac{1}{6}\right)f(-1) + \left(\frac{1}{2}\right)f(0) - \left(\frac{1}{2}\right)f(1) \right. \\
&\quad \left. + \left(\frac{1}{6}\right)f(2) \right]x^3.
\end{aligned} \quad (27)$$

Rearranging the previous equation,

$$\begin{aligned}
P(x) &= f(-1) \left[-\left(\frac{1}{3}\right)x + \left(\frac{1}{2}\right)x^2 - \left(\frac{1}{6}\right)x^3 \right] \\
&+ f(0) \left[1 - \left(\frac{1}{2}\right)x - x^2 + \left(\frac{1}{2}\right)x^3 \right] \\
&+ f(1) \left[x + \left(\frac{1}{2}\right)x^2 - \left(\frac{1}{2}\right)x^3 \right] \\
&+ f(2) \left[-\left(\frac{1}{6}\right)x + \left(\frac{1}{6}\right)x^3 \right].
\end{aligned} \quad (28)$$

Comparing with Eq. (20), we have

$$\begin{aligned}
h_1(x) &= (1-x) \left(1 + \left(\frac{1}{2}\right)x - \left(\frac{1}{2}\right)x^2 \right), \\
h_2(x) &= (1-x)(2-x) \left(\left(\frac{1}{2}\right) - \left(\frac{1}{6}\right)x^3 \right).
\end{aligned} \quad (29)$$

We consider now a generalized derivation of a group of cubic interpolators that includes the TRW cubic as a special case. We mentioned before that the TRW cubic results in an interpolating function that has a continuous first derivative at the sampling points. We show here that this first derivative is a fixed ratio of the slope of the line joining the two adjacent sampling points. If this generalized condition is satisfied at any sampling point, it will be true for any other sampling point. Therefore, we consider the first derivative at the sampling point x_0 . This derivative should equal a fixed ratio α of the difference $f(x_1) - f(x_{-1})$. Thus,

$$\left. \frac{dP(x)}{dx} \right|_0 = \alpha(f(1) - f(-1)). \quad (30)$$

nopqrstuvwxyz
Q=1 bit
Spartan Medium 6 pt

nopqrstuvwxyz
Q=2 bits
Spartan Medium 6 pt

nopqrstuvwxyz
Q=4 bits
Spartan Medium 6 pt

nopqrstuvwxyz
Q=6 bits
Spartan Medium 6 pt

nopqrstuvwxyz
Q=8 bits
Spartan Medium 6 pt

Figure 20 Effects of quantization using bi-linear interpolation.

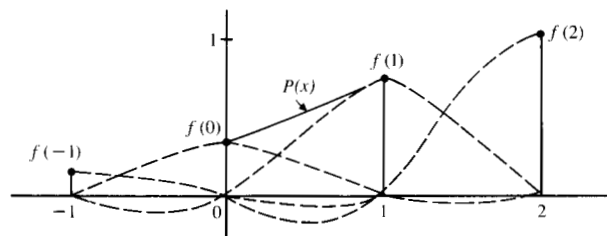


Figure 21 Illustration for cubic convolution.

By substituting the values of $P(x)$, $h_1(x)$, and $h_2(x)$ from Eqs. (20), (23), and (24), Eq. (30) is reduced to

$$\begin{aligned}
f(-1)(-C+D) + f(0)(A-1) + f(1)(A+B+1) \\
+ f(2)(-C-2D) = \alpha(f(1) - f(-1)).
\end{aligned} \quad (31)$$

Since this is true for all values of $f(x)$, the following conditions should be satisfied:

$$\begin{aligned}
 \alpha &= C + D, \\
 0 &= A - 1, \\
 \alpha &= A + B + 1, \\
 0 &= C + 2D.
 \end{aligned}
 \tag{32}$$

Solving in A , B , C , and D , we have

$$\begin{aligned}
 A &= 1, \\
 B &= \alpha - 2, \\
 C &= 2\alpha, \\
 D &= -\alpha.
 \end{aligned}
 \tag{33}$$

If we choose $\alpha = 1$, we then have the TRW cubic interpolator. It is interesting to note that the slope of the interpolated function at any sampling point is twice the slope of a line connecting the two adjacent sampling points. Equality of the two slopes can be achieved if we choose $\alpha = 1/2$; this other case was suggested in [9]. However, it appears that the TRW cubic has better performance in the case of binary pictures. A different derivative of the TRW interpolator is given in [10].

Since all parameters in Eq. (31) are single-valued functions of α , we can change the value of α to optimize the performance for any given signal model. This idea needs further study.

References

1. K. Y. Wong and B. R. Schatz, "Method for Improving Print Quality of Coarse Scan/Fine Print Character Reproduction," U.S. Patent 4,124,870, November 7, 1978.
2. R. W. Schafer and L. R. Rabiner, "A Digital Signal Processing Approach to Interpolation," *Proc. IEEE* **61**, 692-702 (June 1973).
3. E. K. Blum, *Numerical Analysis and Computation*, Addison-Wesley Publishing Co., Reading, MA, 1972.
4. P. Stucki, "Image Processing for Document Production," *Advances in Digital Image Processing*, P. Stucki, Ed., Plenum Press, New York, 1979, pp. 177-201.
5. S. S. Rifman and D. M. McKinnon, "Evaluation of Digital Correction Techniques for ERTS Images," *TRW Report 20634-6003-TU-00*, TRW Systems Group, Redondo Beach, CA 90278, March 1974.
6. W. K. Pratt, *Digital Image Processing*, John Wiley & Sons, Inc., New York, 1978, pp. 111-120.
7. J. V. Uspensky, *Theory of Equations*, McGraw-Hill Book Co., Inc., New York, 1948.
8. I. E. Abdou and K. Y. Wong, "Analysis of Linear Interpolation Schemes for Bi-Level Image Applications," *Research Report RJ2952*, IBM San Jose Research Laboratory, San Jose, CA, October 3, 1980.
9. G. T. Herman and S. W. Rowland, "A Comparative Study of the Use of Linear and Modified Cubic Spline Interpolation for Image Reconstruction," *IEEE Trans. Nuclear Science* **NS-26**, 2879-2894 (April 1979).
10. R. Bernstein, "Digital Image Processing of Earth Observation Sensor Data," *IBM J. Res. Develop.* **20**, 40-57 (January 1976).

Received May 3, 1982; revised July 8, 1982

Ikram Abdou *Aydin Controls, 414 Commerce Drive, Fort Washington, Pennsylvania 19034.* Dr. Abdou received the B.S. (with honors) and the M.S. in electrical engineering from Cairo University, Egypt, in 1970 and 1973, and the Ph.D. from the University of Southern California, Los Angeles, in 1978. He joined Cairo University in 1970, first as a teaching assistant and then as an assistant lecturer. From 1975 until 1978, he was a research assistant with the Image Processing Institute of the University of Southern California. From 1978 until 1980, he was a postdoctoral fellow with the IBM Research laboratory, San Jose, California. He is currently a senior systems engineer with Aydin Controls. Dr. Abdou is a member of the Institute of Electrical and Electronics Engineers and Sigma Xi.

Kwan Y. Wong *IBM Research Division, 5600 Cottle Road, San Jose, California 95193.*

(See page 656 for biography.)