Filters involving derivatives with application to reconstruction from scanned halftone images

Forchhammer, Søren; Jensen, Kim S.

Published in:
IEEE Transactions on Image Processing

Link to article, DOI:
10.1109/83.370673

Publication date:
1995

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Filters Involving Derivatives with Application to Reconstruction from Scanned Halftone Images

Søren Forchhammer and Kim S. Jensen

Abstract—This paper presents a method for designing finite impulse response (FIR) filters for samples of a 2-D signal, e.g., an image, and its gradient. The filters, which are called blended filters, are decomposable in three filters, each separable in 1-D filters on subsets of the data set.

Optimality in the minimum mean square error sense (MMSE) of blended filtering is shown for signals with separable auto-correlation function. Relations between correlation functions for signals and their gradients are derived. Blended filters may be composed from FIR Wiener filters using these relations. Simple blended filters are developed and applied to the problem of gray value image reconstruction from bilevel (scanned) clustered-dot halftone images, which is an application useful in the graphic arts. Reconstruction results are given, showing that reconstruction with higher resolution than the halftone grid is achievable with blended filters.

I. INTRODUCTION

This paper presents and treats a general method for designing finite impulse response (FIR) filters for samples of a signal and its gradient in two dimensions. The filters, which are called blended filters, are composed of three filters, each separable in 1-D filters on subsets of the sample signal and its gradient. This enables the use of filter techniques for 1-D data in the design of blended filters.

Blended filters are applied to the problem of reconstruction from (scanned) halftone images. Halftoning of images is used in the graphic arts to render gray value images as bilevel images. The bilevel images are basically composed of halftone dots, with (relative) areas reflecting the gray values at the specific positions. Besides this gray value information, the shape and (relative) position of the dots might give additional information correlating to the gradient of the original gray value image.

This additional information at halftone subdot level has been used in previous work on halftones. For data compression purposes, methods to determine the area coverage of the halftone (sub)dots have been devised [1], [2]. To utilize the (sub)dot information for image reconstruction, subdot areas were used to estimate gradient values. The reconstruction was thereafter achieved using simple polynomial interpolation [3].

In this paper, emphasis is put on filtering aspects of using gradient samples. Blended filters that are a more general class of filters for samples of a signal and its gradient are defined, and properties of these filters are investigated.

In Section II, blended filters are defined, and a sampling theorem for a specific bandlimitation in 2-D is given. Section III treats the problem of designing blended filters with finite impulse response (FIR). Results are given that may be used for composing blended filters from FIR Wiener filters. The case of a signal with separable autocorrelation function is treated in detail, and conditions under which blended filters are optimal are given. Application to reconstruction from halftone images is devised in Section IV, and specific reconstruction filters are developed. In Section V, numerical values of reconstruction errors for different blended filters on a test image are given, and images reconstructed from a scanned halftone image are presented.

II. BLENDED FILTERS

This section presents a fast method for designing and implementing filters for samples of a signal \( f(x, y) \), e.g., an image, and its gradient \( \nabla f(x, y) = (f_x(x, y), f_y(x, y)) \). The samples are organized in a regular grid with integer coordinates \((m, n)\) in the \((x, y)\)-coordinate system. The samples \( f(m, n) \) of the signal are also referred to as amplitude samples to distinguish them from the gradient samples.

A. Definition of Blended Filters

Filters in 2-D are often realized as separable filters that are decomposable in two 1-D filters to speed up the implementation. For samples of the function \( f(m, n) \) and gradient \((f_x(m, n); f_y(m, n))\), a separable filter for each of the three components may be used. A closely related possibility is to use a blended filter, which is also defined using 1-D filters.

Definition: A blended filter with samples of a signal \( f(m, n) \) and its gradient \((f_x(m, n); f_y(m, n))\) as input has the output

\[
\hat{f}(x, y) = f_x(x, y) + f_y(x, y) - f(x, y)
\]  

(1)

where \( f_x, f_y, \) and \( f \) are intermediate functions given by

\[
f_x(x, y) = f(m, n) * h_{f_x}(x)
\]  

(2a)

\[
f_y(x, y) = f(m, n) * h_{f_y}(y)
\]  

(2b)

\[
f(x, y) = (f(m, n) * h_{f_x}(x)) * h_{f_y}(y) + (f(m, n) * h_{f_y}(y)) * h_{f_x}(x)
\]  

(2c)
The blended filters are separable for $f(z)$ and involve derivative samples and the amplitude samples.

A blended filter may yield a reconstruction in two dimensions from samples of a signal and its gradient in 2-D. The output is obtained by a linear combination of three intermediate reconstruction functions $f_b(x, y) + f_c(x, y) - f_o(x, y)$ (see Fig. 1). The intermediate reconstruction $f_b(x, y)$ is obtained from samples of the signal and its derivative in the $x$ direction ($f(y)$ direction) in a separable manner.

The blended filter may be divided in two steps. The first step makes use of the derivative samples by calculating (1-D) reconstructions $f_b(x, n)$ along the sampling grid lines in the $x$ direction with $y$ coordinate $n$ from samples of the signal $f(m, n)$ and the derivative $f_o(m, n)$ along the specific grid line. $f_c(m, y)$ is calculated by likewise switching the $x$ and $y$ directions. The second step reconstructs $f(x, y)$ by blending the reconstructions along the grid lines $f_b(x, n)$ and $f_c(m, y)$ with the amplitude samples $f(m, n)$ (see (1) and (2)) in the following way. The filter $f_b(x, y)$ with a 1-D impulse response in the $y$ direction is applied to the reconstructed values along the grid lines in the $x$ direction $f_b(x, n)$, to give $f_b(x, y)$. $f_c(x, y)$ is obtained in the same manner applying $h_a(x)$ to $f_c(x, m)$. $f_o(x, y)$ is obtained in the samples of the signal only, using a separable filter. The components of this filter $h_a(x)$ and $h_b(y)$ are also used when calculating $f_b(x, y)$ and $f_c(x, y)$, respectively (see (2b) and (2c)). $f$ is finally obtained from $f_b$, $f_c$, and $f_o$ (see (1)).

In the blended filter (see (1)-(2)), the intermediate reconstruction obtained from amplitude samples $f_a(x, y)$ could be incorporated in $f_b(x, y)$ and $f_c(x, y)$ with the same result. The formulation in (1) was chosen because it conveys the concept of subtracting the amplitude sample reconstruction $f_a(x, y)$ from the sum of two other reconstructions, the latter two involving derivative samples and the amplitude samples.

The second step of the blended filter (obtaining $f(x, y)$ from the discrete amplitude samples $f(m, n)$ and intermediate values along the grid lines $f_b(x, n)$, and $f_c(m, y)$) has the structure of a blended interpolant as described in [4]. If the impulse responses of the 1-D filters at grid points are zero except for $h_{b}(0) = h_{c}(0) = h_{o}(0) = 1$, then the output of the blended filter along the grid lines are given by $f_b(x, n)$ and $f_c(m, y)$. In this special case, the second step of the blended filter takes the form of a blended interpolant as described in [4]. In this case, the functions of the second step comply with the requirements $f(m, n) = f_b(m, n) + f_o(m, n) + f_c(m, n)$ and $f(m, y) = f_c(m, y)$ of a blended interpolant. Generally, blended interpolants [4] differ from blended filters by being defined by functional values along the grid lines, e.g., $f_b(x, n)$ and $f_c(m, y)$ and not by discrete (gradient) samples. Another difference is the requirements of the interpolation functions of the blended filter listed above. The structure in which the filter (interpolation) applied to $f(m, n)$ is composed of the two applied to $f_b(x, n)$ and $f_c(m, y)$ as in (2a)-(2c) and the blending in (1) is common to the two blending techniques.

In the special case above, the output values along the grid lines only require one amplitude-derivative filter with 1-D impulse response. This means that reconstruction along grid lines becomes very simple. The terms amplitude-derivative and amplitude-gradient filter refer to filters operating on input samples of the amplitude and one or two derivatives, respectively.

With symmetry in $x$ and $y$, we have a symmetric blended filter. In this case, the blended filter is described by two 1-D filters: one amplitude-derivative filter and one amplitude-only filter.

Incorporating $f_a$ in the expression(s) of $f_b$ and/or $f_c$, the filter uses (up to) six 1-D convolutions for one output value, which is the same as for the separable filter.

B. Amplitude-Gradient Sampling Theorem for Blended Filters

Classical results on sampling (in 1-D) may be combined in the blended filter structure to derive a sampling theorem for amplitude-gradient data in a simple manner [3].

The classical result for reconstruction of a function $f(t)$ in 1-D, bandlimited to the interval $(-\omega_a, \omega_a)$, i.e., with Fourier transform $F(\omega)$ zero outside this interval from samples with a regular spacing $T = \pi/\omega_a$ is [5]

$$f(t) = \sum_{n=-\infty}^{\infty} \frac{\sin(\omega_0(t-nT)) f(nT)}{\omega_0 (t-nT)}.$$  (3)

For $f(t)$ bandlimited to the interval $(-\omega_a, \omega_a)$ and sampled along with the derivative $f'(t)$ in 1-D with a regular spacing of $2T$, the reconstruction formula is [6]

$$f(t) = \frac{\sin^2(\omega_0 t/2)}{\omega_0^2} \sum_{n=-\infty}^{\infty} \left[ \frac{f(2nT)}{(t-2nT)^2} + \frac{f'(2nT)}{(t-2nT)^2} \right].$$  (4)

The reconstruction formula (4) resolves the (2 times) spectral overlap, due to the reduced (halved) sampling rate of each of the components $f$ and $f'$. The assumptions that the Fourier transform of $f$ exists and is bandlimited ensures that the partial derivatives and their Fourier transforms exist.
The ideal reconstruction filters above (3) and (4) may be used as the 1-D filters of a symmetric blended filter. The reconstruction filter for amplitude-derivative samples in 1-D (see (4)) is used for samples of the function and the partial derivative in the \( x/y \) direction to obtain \( f_{\theta_0}/(f_{\theta_0}) \) along the grid lines in (2b) and (2c). The reconstruction filter for samples of only the function in (3) is used as \( h_{\theta_0} \) and \( h_{\theta_0} \) in (2a)-(2c). This way, exact reconstruction for a 2-D signal with bandlimitation specified as shown in Fig. 2 is obtained [3].

Petersen and Middleton [7] gave a detailed treatment of reconstruction from the samples of amplitude and gradient of an \( n \)-dimensional stochastic field. They showed that exact reconstruction (in the zero mean square error sense) is possible if the spectral images induced by sampling do not overlap more than \( (n+1) \) times. The blended filter obtained above using the filters (see (3) and (4)) resolves the 3 times spectral overlap for the bandlimitation specified by Fig. 2 in a simple manner [3], [8].

If the ideal lowpass filter of (3) is exchanged with an ideal bandpass filter with passband \( (\omega_1/2, \omega_1) \), the corresponding symmetric blended filter yields exact reconstruction for the bandpass limitation with \( \omega_1/2 < \max \{ \omega_1, \omega_2 \} < \omega_1 \) [8].

III. DESIGNING FIR BLENDED FILTERS

In this section, the problem of designing FIR blended filters for practical reconstruction using samples of the 2-D signal and its gradient is considered.

In image processing, FIR filters are often preferred because they enable the design of linear/zero phase filters. This attribute is important [9] because it prevents frequencies in phase at an edge from being forced out of phase, thus distorting the edge and blurring the image.

A number of techniques for designing 1-D and 2-D filters [10] may be extended to the case of 2-D amplitude-gradient filtering. Examples of this are windowing, polynomial interpolation, Wiener filters, and frequency sampling. One reason for using a blended filter is simple design and implementation. Using a blended filter, the problem of designing the filters may immediately be reduced to a problem of designing 1-D (amplitude-derivative) filters. An example of this, using the windowing technique, is to truncate the ideal impulse responses of (3) and (4). However, generalizing a 1-D technique that is optimal for amplitude samples in some sense does not imply that the corresponding 2-D blended filter is optimal. Truncating the ideal impulse response(s) does not give a min. mean square error (MMSE) approximation of the impulse response for a 2-D blended filter as it does for 1-D (amplitude-only) filter.

In the following, the use of polynomial interpolation for designing blended filters is briefly described. Thereafter, the use of FIR Wiener filters is treated, emphasizing signals with separable autocorrelation function. The optimality of blended Wiener filters is also examined.

A. Polynomial Interpolation

Lagrange interpolation is often used in image processing, especially in the graphic arts [9]. Lagrange interpolation uses the polynomial of least degree matching a specified number of points, Hermite interpolation [4] generalizes this in 1-D to the case of samples of the function and its derivatives up to some degree (which may vary from point to point), again using the polynomial of least degree that matches the sample data. The generalization to functions of higher dimensions is often called (Hermite-)Birkhoff interpolation [11]. In this case, the polynomial space [11] has a dimension coinciding with the number of polynomial data, i.e., sample values within the current window.

Filters corresponding to piecewise Lagrange and Hermite interpolation, with impulse responses derived from the polynomials [12], may be used for blended filters (see (1) and (2)). Hermite interpolation may be used piecewise to obtain \( f_{\theta_0}(x, n) \) and \( f_{\theta_0}(m, y) \) in the first step (see (2b) and (2c)) of the filter. Applying the filters \( h_{\theta_0}(y) \) and \( h_{\theta_0}(x) \), corresponding to a piecewise Lagrange interpolation, in the second step is equivalent to a blended interpolation of \( f(m, n) \), \( f_{\theta_0}(x, n) \), and \( f_{\theta_0}(m, y) \). A simple example is for four neighboring grid points to use linear interpolation for the filters \( h_{\theta_0}(y) \) and \( h_{\theta_0}(x) \), corresponding to a bilinearly blended Coons patch [13]. Hermite interpolation may be used for the interpolation along the grid lines to obtain \( f_{\theta_0}(x, n) \) and \( f_{\theta_0}(m, y) \), which, as mentioned previously, are presumed given when using a blended interpolant. Using a rectangular subset of samples for polynomial interpolation, the corresponding blended filter is a (Hermite-)Birkhoff interpolation [14].

For interpolation involving noisy data, it is sometimes advantageous to attenuate the use of (gradient) data [3]. This way, blended filters, which do not match the data set and thus differ from the blended interpolants, are obtained.

B. FIR Wiener Filters for Amplitude-Gradient Samples

In general, having the observations \( n_1, n_2, ..., n_m \), the (FIR) linear estimate of the variable \( u \) is given by

\[
\hat{u} = \sum_{j=1}^{m} h_j n_j.
\]
The best linear estimator with respect to MMSE using m
observations is given by the m well-known equations [15]
\[
\sum_{k=1}^{m} E[y_i y_k] h_k = E[y_i v_j], j \in \{1, m\}.
\] (6)

If the variables are gaussian distributed, the estimator is
optimal. The observations may be of any dimension, and
gradients may also be incorporated. Therefore, the case of
estimation from samples of a 2-D signal, and its gradient is
just one special case. If the observations are from stationary
process(es), the mean values of (6) may be expressed by the
evaluation from samples of a 2-D signal, and its gradient is
gradients may also be incorporated. Therefore, the case of
optimal. The observations may be of any dimension, and
functions of (6). we may describe the derivative
of the signal and its derivatives may be derived using the
process, relations between the spectras of the observations
obtained from the corresponding spectras. The correspondence
is obtained from the impulse response.

In one dimension, we get from the theory of linear systems
the correlations in one dimension, we get from the theory of linear systems
\[
S_x(\omega) = \omega^2 S_x(\omega) \leftrightarrow r_x(t) = -r_x^\prime(t)
\] (7)
\[
S_{xx}(\omega) = -i\omega S_x(\omega) \leftrightarrow r_{xx}(t) = -r_x^\prime(t).
\] (8)

The results from linear systems also apply in 2-D for the partial
derivatives [5], giving the relations of the spectras and thereby
the autocorrelation functions
\[
S_{xx}(\omega_1, \omega_2) = \omega_1^2 S_{f_1}(\omega_1, \omega_2) \leftrightarrow r_{xx}(x, y) = -\delta_{x} r_{x}^\prime(x, y) - \delta_{y} r_{y}^\prime(x, y).
\] (9)

\[
S_{ff}(\omega_1, \omega_2) = \omega_1^2 S_{f_1}(\omega_1, \omega_2) \leftrightarrow r_{ff}(x, y)
\]
\[
= -\delta_{x} r_{x}^\prime(x, y) - \delta_{y} r_{y}^\prime(x, y).
\] (10)

\[
S_{ff}(\omega_1, \omega_2) = \omega_1^2 S_{f_1}(\omega_1, \omega_2) \leftrightarrow r_{ff}(x, y)
\]
\[
= -\delta_{x} r_{x}^\prime(x, y) - \delta_{y} r_{y}^\prime(x, y).
\] (11)

Writing \( f_{xy}(x, y) \) as the output of a linear system with input
\( f_{x}(x, y) \) gives the transfer function
\[ H(\omega_1, \omega_2) = i\omega_y / i\omega_1, \] which gives
\[
S_{ff}(\omega_1, \omega_2) = (i\omega_2 / i\omega_1) S_{f_1}(\omega_1, \omega_2)
\]
\[
= -\delta_{x} r_{x}^\prime(x, y) - \delta_{y} r_{y}^\prime(x, y).
\] (12)

The relations between the correlation functions in one dimen-
sion (see (7) and (8)) were derived in [17] differentiating
the expected values without the use of the transfer functions.
The same way, the correlation results in 2-D (see (9)-(12))
may be derived without the transfer functions and, thereby,
the assumption of bandlimitation. The cross correlation \( r_{ff}(x, y) \)

For \( f \), \( f \), and \( f \) separable, comparing the structure of the
derivatives (1-D) and the partial derivatives (2-D) in (7)-(13),
the following relations are obtained:
\[
\frac{\partial r_{ff}(\tau_1, \tau_2)}{\partial \tau_1} = -\frac{\partial^2 r_{f}(\tau_1, \tau_2)}{\partial \tau_1 \partial \tau_2}
\] (15)

which is the result of the right-hand side of (13).

1) Correlation Functions for Signals with Separable Auto-
correlation: For a signal with separable autocorrelation
function
\[
r_{f}(x, y) = r_{x}(x)r_{y}(y)
\] (16)

the 2-D process \( f \) may be decomposed in 1-D processes in
the \( x \) and \( y \) directions with derivatives \( f \) and \( f \), respectively.
In this case, there are simple relations between the correlation
functions.

For \( r_{f}(x, y) \) separable, comparing the structure of the
derivatives (1-D) and the partial derivatives (2-D) in (7)-(13),
the following relations are obtained:
\[
r_{ff}(x, y) = r_{x}(x)r_{y}(y)
\] (17)
\[
r_{ff}(x, y) = r_{x}(x)r_{y}(y)
\] (18)
\[
r_{f}(x, y) = r_{x}(x)r_{y}(y)
\] (19)
\[
r_{f}(x, y) = r_{x}(x)r_{y}(y)
\] (20)
\[
r_{f}(x, y) = -r_{x}(x)r_{y}(y).
\] (21)

The relations between the correlation functions may be
used when setting up (6), describing the Wiener filter. In the
following section, they are used to derive results on blended
filters for signals with separable autocorrelation.

2) Optimal Blended Filters for Separable Autocorrelation
Functions: In the following, the optimality of blended filters
and the intermediate functions \( f_{x}, f_{y} \), and \( f_{xy} \) is examined.
Equation (6) may be written in matrix form. For stationary
processes, the expected values are described by correlation
values. The matrix of (6) is written with capital bold and
the vectors with lower-case bold. The subscripts refer to
the processes involved and the subset of data used. The 2-D
processes \( f, f_{x} \), and \( f_{y} \) are decomposed in the \( x \) and \( y \)
direction. Let subscript 1 refer to the \( x \)-direction, 2 to the \( y \)-direction. Let subscript \( f_{x} \), \( f_{y} \), and \( f_{xy} \) refer to the process(es) involved.

Writing \( f_{xy}(x, y) \) as the output of a linear system with input
\( f_{x}(x, y) \) gives the transfer function
\[ H(\omega_1, \omega_2) = i\omega_y / i\omega_1, \]
which gives
\[
S_{ff}(\omega_1, \omega_2) = (i\omega_2 / i\omega_1) S_{f_1}(\omega_1, \omega_2)
\]
\[
= -\delta_{x} r_{x}^\prime(x, y) - \delta_{y} r_{y}^\prime(x, y).
\] (12)
The MMSE linear interpolation of \( f(x, y) \) is given by a blended filter \( f(x, y) = f_a(x, y) + f_b(x, y) \) for a rectangular region of support for each of the components given by \( w_a(x, y) \) and \( w_b(x, y) \) are the MMSE linear solutions given by Lemma 1. \( f_a(x, y) \) is the separable MMSE linear solution (22) for the samples of \( f(x, y) \).

Proof: See Appendix A for proof.

The autocorrelation functions given by \( r_s(x, y) = \exp(-\alpha(x^2 + y^2)) \), where \( a \) and \( c \) are constants, are an example of autocorrelation functions for which the Lemma and the Theorem may be applied. \( r_s \) describes the only rotation invariant separable functions. For bandlimited signals, the results may be applied as the derivatives of the signal and the autocorrelation exist, thereby satisfying the assumptions.

3) Composite Wiener Filtering: Images are, in general, not stationary. A composite model is a better description. In this case, the image (model) is composed of a number of statistically distinct stationary objects belonging to a set of \( K \) classes. This model leads to a simple extension of the FIR Wiener filter, which adapts to the image locally.

Lebedev and Mirkin [19] introduced a filter weighting the FIR Wiener filter \( f_s(x, y) \) of each class \( \theta \). The weighting functions

\[
p(\theta \mid w(x, y))
\]

are the conditional probability for \( f(x, y) \) that belong to the class \( \theta \) given the data within the window of the filter \( w(x, y) \). The composite estimate \( f(x, y) \) given by weighting the Wiener estimates of each class is

\[
f(x, y) = \sum_{\theta=1}^{K} p(\theta \mid w(x, y)) f_s(x, y).
\]

Under the assumptions of Gaussian distributions of the image and additive and independent noise within each class, over the filter window, Lebedev and Mirkin showed that this estimate is optimal (MMSE) for smoothing an image. In the next section, the estimate (24) is modified by letting \( M(x, y) \) be a vector function that operates on the data within a window in the neighborhood of \( (x, y) \). This will be called a composite Wiener filter.

IV. APPLICATION TO RECONSTRUCTION FROM HALFTONES

In this section, application of amplitude-gradient filtering to the problem of reconstruction from scanned halftone images is described. A scanner delivers a (high-resolution) bilevel image representing the halftone. This is to be converted into a (digital) gray value image at a (lower) resolution, which preferably is higher than that of the halftone grid. One interesting point is the acquisition of gradient samples with information not contained by the amplitude samples.

Basically, halftone images are bilevel images where dot areas represent gray values, but there may be more information available than just the area of the halftone dot. That is, the halftone image may convey more information than a sampling at the halftone screen resolution. To design high-resolution reconstruction schemes for halftone images, it is therefore necessary to take a closer look at the screening method used.

In conventional halftoning, the screen function is added to the gray value image and thresholded in a photomechanical process. Threshold screening (or electronic halftoning [20]) is a digital equivalent of conventional screening. The gray

\[
\text{Fig. 3. Rectangular regions of support for blended filters interpolating \( f \) at \( (0) \). Left: Amplitude (\( \times \)) and derivative \( f_x = \pm 1 \) samples are used as in the lemma. Right: Amplitude (\( \times \)) and derivative \( f_x = \pm 1 \) samples are used as in the theorem.}
\]
value image function \( f(x, y) \) is thresholded with the periodic threshold (or screen) function to create the resulting bilevel image. Considering clustered dots, the transitions in the bilevel halftone image represent the crossings of the image function and a threshold (or screen) function (Fig. 4). This way, the halftone dots are shifted relative to the grid point according to the gradient of the image function across the dot. Clustered dot halftoning, as opposed to split dot dithering, is used in offset printing in the graphic arts industry. Reconstruction from halftone images may be used 1) to input (old) halftone material, 2) for data compression, and 3) for offset to gravure conversion.

The problem of reconstruction from halftones may be described as a reversal of the nonlinear screening process. Using a (nonlinear) estimate of the gradient (25) transforms the problem into a problem of reconstruction from (amplitude-gradient) samples in a regular grid. This makes it easy to apply local methods. Operating directly on the gradient may also have advantages as it relates to the edges (or high frequencies), which are important in image processing.

Other approaches to the reconstruction problem include filtering in the frequency domain, reconstruction from nonuniform samples [21], reconstruction from level crossings [22], and projections onto convex sets [23]. These techniques are more global methods, and the last three are relatively complex and vulnerable to the inaccuracies in the estimation of the phase of the halftone screen frequency. However, they have the potentials of yielding reconstructions with even higher resolution.

A. Acquisition of Gradient Samples

In previous work on scanned halftone images [1], [2] directed toward data compression, the halftone dots have been located and divided in four triangular regions. The areas of these subdots are determined. The region connected with a whole dot is called a cell (Fig. 5). For data compression purposes, the (sub)dot areas may be coded to represent the image. Here, as in [3], the subdot areas are used to obtain estimates of the gradient, and the cell dot areas are converted into amplitude samples. The triangle and cell values are first corrected to reduce quantization effects [14].

The derivatives \( f_x \) and \( f_y \) at a cell sample point can be expressed as a function of the sample values in the four triangles forming the cell (white or black), assuming that the derivatives are constant across the dot area. Let \( s_1, s_2, s_3, \) and \( s_4 \) be the values of the four triangles of a cell (Fig. 5). For clustered halftone dots with a diamond dot shape, geometric calculations give the following estimate of the partial derivatives in the cell center [3]:

\[
\begin{align*}
\frac{f_x}{f_y} &= \frac{2\sqrt{2}(s_2 - s_4)}{\sqrt{(s_1 + s_2 + s_3 + s_4)}} \\
&= \frac{2\sqrt{2}(s_3 - s_1)}{\sqrt{(s_1 + s_2 + s_3 + s_4)}}, \quad s_i \in [0; 1].
\end{align*}
\]

The idea of acquiring amplitude and gradient samples may be used for other types of halftone dots (including dispersed dots) as long as the gradient of the original image correlates to the bilevel halftone image. The gradient expression should ideally reflect the halftoning method used.

B. Blended Filters for Reconstruction from Halftones

Having acquired both amplitude and gradient samples, the methods for designing blended filters may be applied to these data. As mentioned previously, designing symmetric blended filters (see (1) and (2)) for 2-D data only requires the design of two 1-D filters. The reconstructions rendered at the end of this section have four gray value samples per halftone dot (before being screened again for reproduction). Using four gray value samples per halftone dot is a rule of thumb in the graphic arts. It is also theoretically sufficient for maintaining the information of a signal bandlimited as depicted on Fig. 2. A halftone image has two interlaced grids: one with white dots and one with black dots at the grid points (Fig. 5). At four samples per dot, using the samples of both the black and the white grid, only one 1-D amplitude-derivative filter is required to interpolate the intermediate samples (Fig. 5). This means a discrete 1-D filter that doubles the resolution along the black and/or white grid lines may be used. To keep the filter simple, the intermediate samples are found from the intermediate functions \( f_0 \) and \( f_1 \) of the blended filter along the lines of the black grid, whereas only the amplitude samples are used from the white grid. The design is hereafter limited to the 1-D amplitude-derivative filters doubling the resolution along the grid lines of this setup.
and different design methods are given in Fig. 6 normalized to

The polynomial (Hermite) filter (see Section III) and the filters obtained by truncating the ideal impulse responses of the filters of Section II (see (3) and (4)) are data independent and straightforward to calculate.

The Wiener filters and the composite filters of Section III require a model and/or statistical measurement of the data. The stationarity of image data may be improved by subtracting the local mean and normalizing by the local variance [16], [24]. A model with a (slowly) varying mean value added to a zero-mean signal is used. The model imposed also assumes symmetry giving symmetric filters. The local mean is subtracted, and here, the local variance may be used to control the weighting functions of a composite filter. The local mean is calculated within the filter window to simplify the filter. The Wiener filters used are based on measured mean values of the products of the observations (6) (with mean value subtracted) within the filter window. The resulting filter output used is the result of the Wiener filter for the zero-mean data plus the mean value within the filter window. The resulting filter will be referred to as a (constrained) Wiener filter. This implies that the mean value of the original amplitude values is maintained, giving a nonbiased estimate for the nonzero mean image data. This also implies that the sum of the coefficients for the amplitude samples is one.

To design the composite filter, four classes \( H \) (or 'training sets') of the test image (Fig. 10) are chosen. These are parts of the mug (\( \bigstar \)), the flower leaves (\( \Delta \)), the flower centers (\( \bullet \)), and the calm background region (\( \cdot \)). The (local) variance of the amplitude samples within the window filter and a Beaudet derivative estimate [25] (in the \( x \) or \( y \) direction) in the point to be interpolated are used as the 2-D vector function \( f(x, y) \), which weights the filters of the different classes in (24). The Beaudet operator used is a first-order approximation to the first-order derivative with an eight-cell samples support [14].

Let a \( (m,n) \) prefix denote a filter in 1-D with \( m \) amplitude and \( n \) derivative samples. Filter coefficients resulting from the different design methods are given in Fig. 6 normalized to the doubled sampling rate. The filters are symmetric, and the sum of the coefficients for the amplitude samples equals one for the (constrained) Wiener filters. Therefore, the coefficients \( (h_3, h_1, h_2, h_3) \) and \( (l_1, l_2) \) describe the (4, 2) Wiener filters with \( h_3 = 1/2 - h_2, h_1, h_2 \) and \( h_3 \) are the coefficients of the nearest two and the next two amplitude values, respectively. \( l_1 \) is the numerical value of the coefficient for the two nearest derivative samples. The data used is the threshold screened test image presented in Section V. A Wiener filter is designed for each of the four classes (\( \bigstar, \Delta, \bullet, \cdot \)) into which the image may be divided. The image statistics has also been calculated as a function of the Beaudet derivative estimate and the local variance, respectively. The corresponding Wiener filters are shown connected according to a monotonic relation of these derivative and variance measures. The figure indicates that a composite filter based on the four classes may span the Wiener filters determined according to specific values of the weighting functions. The composite filter is a simple way to design an adaptive filter obtaining the filter from a training set.

C. Transfer Function of Amplitude-Derivative Filter

The effects of the filters may be examined in the frequency domain. Assuming the derivative to be the output of a linear system with the signal as input and the transfer function \( H(w) = i\omega \), a combined transfer function for differentiation and amplitude-derivative filtering may be calculated. Normalized to the doubled sampling rate, the transfer function is

\[
H(w) = 0.5 + h_1 \cos(\omega) + h_2 \cos(3\omega) + l_1 \omega \sin(\omega) \tag{26}
\]

for a (4,2) 1-D symmetric filter doubling the sampling rate along a grid line. This transfer function is shown (Fig. 7) for the (4,2) Wiener filter obtained using the data of the mug. This shows that the derivative data enhances the high frequencies from one half \( f_s/2 \) to the full halftone grid frequency \( f_s \).

If the original signal has frequencies above \( f_s/2 \), aliasing occurs for the amplitude and derivative samples taken separately. Combining these samples, the spectral overlap is (ideally) resolved by (4), i.e., no aliasing occurs if the signal is bandlimited to \( f_s \). Assuming this bandlimitation \( (\omega < \pi) \), the aliasing for the (4,2) filter of (26) may be described by

\[
H'(\omega) = 0.5 + h_1 \cos(\pi - \omega) + h_3 \cos(3(\pi - \omega)) + l_1 \omega \sin(\pi - \omega), 0 \leq \omega < \pi \tag{27}
\]

which expresses the attenuation of the aliasing coming from frequencies originally located at \( \omega \). Fig. 8 shows this aliasing attenuation for the (4,2) Wiener filter of the mug.

V. RESULTS

In this section, the results of applying blended filters to amplitude-gradient samples acquired from halftone images are given. The object is reconstruction of gray value images from bilevel (scanned) halftone images. The reconstruction results are evaluated both objectively (MSE) and subjectively.

Two test images have been used for development and evaluation of the reconstruction filters.

1) The first image is referred to as the scanned test image (Fig. 10). It is screened using conventional contact screening.
Fig. 7. Transfer function up to the screen frequency (\(\pi\)) of filter (solid line) doubling the sampling rate along a grid line. The (4,2) 1-D filter derived from the data of the mug part of the test image is used to interpolate the intermediate points. The transfer function (dotted line) with the derivative samples left out is shown for comparison.

Fig. 8. Attenuation \(H'(\omega)\) of aliasing frequencies originally located at \(\omega\) from 0 to the screen frequency (\(\pi\)) corresponding to half the new rate.

and thereafter scanned as a bilevel image at 640 dots/cm on a SCITEX Raystar giving a 4864 by 6144 pixel bilevel image. The halftone screen is angled at 63° with a distance of 12.3 bilevel pixels between halftone dots. A line code was added after the screening but before the scanning. The scanned test image is the primary test image for visual evaluation of reconstruction results.

2) The second image, which is referred to as the threshold screened test image, is derived from the same photograph as the first but scanned as a gray value image on a RASTEC Pixact scanner at 100 dots/cm. This digital image has been used to create a (continuous) reference image function by using threshold screening together with a 2-D separable Lagrange interpolation of third order (in 1-D). The screen angle is 63° and the dot spacing the same as above, i.e., 12.3 pixels between the halftone dots. The data of this image has been used to develop the filters and measure errors in the reconstructions.

The derivative estimate (in the \(y\) direction by (25)) from the scanned test image is shown (Fig 9). Besides showing noise at the screen frequency, this plot illustrates low-frequency patterns due to small variations in the actual halftone grid together with quantization of the grid estimate. It is important to avoid these artifact patterns in the reconstructions. As seen in the images presented below, the reconstruction filters succeed in this.

A. Error Measurements

As mentioned, the threshold screened test image was generated digitally from a continuous description of a gray value image. This enables measurements of errors as the image function may be calculated at any point. This is used to measure the errors of the cell and the triangle samples that are the input to the derivative estimate (25) and the succeeding reconstruction filter. The errors of the reconstructed images are also measured.

Table I illustrates the quality of the cell and triangle samples after the transformation of dot areas into amplitude values (see Section IV). The mean square error (MSE) and peak signal-
Fig. 10. Scanned bi-level test image.

Table I: Error Statistics (MSE and PSNR) for Cell and Triangle Amplitude Samples Measured Over the Threshold Screened Test Image

<table>
<thead>
<tr>
<th>Amplitude values</th>
<th>MSE</th>
<th>PSNR dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>cell samples</td>
<td>19.9</td>
<td>35.2</td>
</tr>
<tr>
<td>triangle samples</td>
<td>53.5</td>
<td>30.9</td>
</tr>
</tbody>
</table>

Table II: Error Statistics (MSE and PSNR) for Reconstructions of the Test Image Using Lagrange and (Composite) Wiener Filters

<table>
<thead>
<tr>
<th>Filter</th>
<th>MSE</th>
<th>PSNR dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4,0) Lagrange</td>
<td>32.9</td>
<td>33.0</td>
</tr>
<tr>
<td>(6,0) Wiener</td>
<td>24.8</td>
<td>34.2</td>
</tr>
<tr>
<td>(4,2) Wiener</td>
<td>19.5</td>
<td>35.3</td>
</tr>
<tr>
<td>(4,2) Composite</td>
<td>17.7</td>
<td>35.7</td>
</tr>
</tbody>
</table>

Fig. 11. Reconstruction from scanned test image shown on Fig. 10 using the (4,2) Wiener filter.

To test the methods on real data and for subjective evaluation, the scanned test image has been used for reconstruction using the filters designed for the threshold screened test image. The reconstruction result using a (4,2) Wiener filter is presented in Fig. 11.

The (composite) Wiener filters designed in this paper are obtained from the digitally screened test image. The filters will therefore reflect the screening method and the masking method used in the processing.

B. Image Results

To test the methods on real data and for subjective evaluation, the scanned test image has been used for reconstruction using the filters designed for the threshold screened test image. The reconstruction result using a (4,2) Wiener filter is presented in Fig. 11.

Comparing the Wiener filter reconstruction (Fig. 11) with the scanned bilevel image (Fig. 10), the reconstruction is not quite as sharp but is still a good reconstruction. The reduced sharpness is mainly seen in the line code, the letters...
Fig. 12. Enlargements of part of reconstructed images. From top to bottom: (a) Lagrange polynomial interpolated image \( f_n \); (b) Hermite polynomial interpolated image \( f_n \) using the samples of the derivate in the \( x \) direction; (c) image reconstructed using a polynomial blending filter \( f \). The improvement is mainly seen in the background of the mug.

(which have not been screened; see also Fig. 13) in the detailed parts of the cup, and in the flower centers. The Wiener reconstructions give images with very few and hardly visible artifacts and better sharpness than, e.g., a Lagrange interpolated image from amplitude samples [14]. The images obtained by composite Wiener filtering are somewhat sharper than the simple Wiener reconstructions, but some overshoot appears at sharp edges [14]. This artifact could be reduced, however.

The enlarged part of the image in Fig. 12 illustrates the effect of using gradient data and blended filters. The difference is mainly seen at edges and in detailed areas, e.g., the textured background of the mug.

Fig. 13 illustrates that a reconstructed image may be used as input to a system for further processing, e.g., screening with a different halftone screen in the graphic arts. A very close look at the scanned halftone image (Fig. 10) and the reconstruction (Fig. 11) will also reveal the same difference in the angle of the halftone grid.

VI. CONCLUSION

Filters for interpolation from samples of an image function and its gradient have been treated. A blended filter technique that composes the 2-D filter of 1-D filters has been introduced. For signals with separable autocorrelation function, blended filters have been shown to be (MMSE) optimal for linear interpolation over a rectangular region of support. The
filter is decomposed in 1-D MMSE linear filters. Other filter techniques in 1-D may also be used to design blended filters.

Blended filters have been applied to the problem of image reconstruction from (scanned) bilevel halftone images. An application useful in the graphic arts industry. Local mean values were subtracted and 1-D MMSE Wiener filters were designed. Local variance and derivative estimates were used to control an adaptive composite Wiener filter based on Wiener, and polynomial blended filters all gave good results ranked in the given order. The derivative data are used in 1-D filters of small sizes with four amplitude samples and two derivative samples. The resulting reconstructions based on local data have high resolution (higher than that of the halftone grid).

APPENDIX A

PROOF OF THEOREM

To prove the lemma and the theorem, some matrix results and notation are used. The following properties [16] of the Kronecker product are used:

\[(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}\]  
\[(A \otimes B) \cdot (C \otimes D) = A \cdot C \otimes B \cdot D\]  
\[(A \otimes B)^T = A^T \otimes B^T\]

where \(T\) denotes transposition.

Equation (6) may be written in matrix form. Let subscripts \(x'\) and \(y'\) of vectors and matrices refer to the derivatives in the 1-D processes in the respective directions. The subscripts \(a, b,\) and \(c\) refer to the subsets of data involved in the respective intermediate functions \(f_a, f_b,\) or \(f_c\) of the blended filter (see (1) and (2)). An additional subscripting of \(a, b,\) and \(c\) with \(x\) or \(y\) denotes decomposition in the given direction. The matrices are organized by letting the indices \(j\) and \(k\) of (6) run through the samples in the order \(f, f_x,\) and \(f_y.\) Within each set of data the order is row by row over the filter window. This organization of data is equivalent to forming a 1-D set for each output of the filter by concatenating the data in the order \(f, f_x,\) and \(f_y\) and within each component using row ordering.

For a separable autocorrelation function \(r_f, (16)-(21)\) are valid for the matrices and vectors of the equations substituting multiplication with the Kronecker product. The decomposition of the 2-D Wiener filter \(h_f\) of (22) in the two 1-D Wiener filters \(h_x\) and \(h_y\) uses (A1) and (A2).

Proof of Lemma 1: The 2-D solution to be decomposed is given by

\[h_b = (R_{by})^{-1}r_{by}.\]  
(A4)

The Wiener filter in 1-D involving samples of \(f\) and \(f_x\) is given by

\[h_{fx} = (R_{fx})^{-1}r_{fx}\]  
(A5)

\[R_{fx} = \begin{bmatrix} R_x & R_{xx'} \\ R_{xx'}^T & R_{xx} \end{bmatrix}\]  
(A6a)

\[h_{fx} = [h_{fx}, h_{fx}]\]  
(A6b)

where \([\cdot]\) denotes concatenation. The Wiener filter in the \(y\) direction is given by

\[h_{by} = (R_{by})^{-1}r_{by}.\]  
(A7)

Combining \(h_x\) and \(h_y\) by direct matrix multiplication, using (16), (17), (19), (A2), (A3), and \(R_b = R_{by}^T,\) we get

\[h_b = h_x \otimes h_y = ((R_{bx})^{-1}r_{bx}) \otimes ((R_{by})^{-1}r_{by})\]  
\[= ((R_{bx})^{-1} \otimes (R_{by})^{-1})(r_{bx} \otimes r_{by})\]  
\[= \left[ \begin{array}{cc} R_x & R_{xx'} \\ R_{xx'}^T & R_{xx} \end{array} \right]^{-1} (r_{bx} \otimes r_{by})\]  
\[= \left[ \begin{array}{cc} R_f & R_{ffx} \\ R_{ffx}^T & R_{ffy} \end{array} \right]^{-1} r_{by}\]  
(A8)

The last line is the Wiener filter (A4). Given samples of the function \(f\) and the other derivative \(f_y, the results are found by symmetry (in \(x\) and \(y)).\)

Proof of Theorem 1: The Wiener reconstruction filter \((h)\) may be written in the form

\[h^T = [(h_{fx} + h_{fy} - h_b)^T h_{fx}^T, h_{fy}^T]\]  
(A9)

where \(h_{fx}^T = \left[h_{fx}, h_{fy}\right]^T\) and \(h_{fy}^T = \left[h_{fy}, h_{fy}\right]^T,\) if it is a blended filter. Furthermore, \(h_b\) should be the solution to (6), which may be written as

\[\begin{bmatrix} R_f & R_{ffx} & R_{ffy} \\ R_{ffx}^T & R_{ffx} & R_{ffy} \\ R_{ffy} & R_{ffy} & R_{ffy} \end{bmatrix} \begin{bmatrix} h_{fx} + h_{fy} - h_b \\ h_{fx} \\ R_{ffx} \\ h_{fy} \\ h_{fy} \end{bmatrix} = \begin{bmatrix} r_f \\ R_{ffx} \end{bmatrix}\]  
(A10)

The equations involving \(r_f\) on the right-hand side are satisfied by inserting each of the solutions \(h_b\) given by (22) and \(h_b\) and \(h_b\) from Lemma 1. These equations are, therefore, also satisfied by \(h_b + h_c - h_b.\) For the equations involving \(r_{ffx},\) on the right-hand side and subtracting the solution of Lemma 7.1, \(h_b, i.e.

\[R_{ffx}^T (h_{fx} + h_{fy} - h_b) + R_{ffx} h_{ffx} = 0.\]  
(A12)

To prove that this holds, the following identities valid over the specified region of support of the filters are used with \(h_{fy}^T = \left[h_{fx}, h_{fy}\right]^T:\n
\[h_b = h_{fa} \otimes h_{fa}\]  
(from 22)

\[h_{fx} = h_{fa} \otimes h_{fx}\]  
(using Lemma 1 in \(y)\)

\[h_{fy} = h_{fa} \otimes h_{fx}\]  
(using Lemma 1 in \(y)\)

\[R_{ffx} = R_{fx} \otimes R_{yy}\]  
(by 17)

\[R_{ffx} = -R_{xx} \otimes R_{yy}\]  
(by 21)

Using (A1) and (A2), we get

\[R_{ffx}^T (h_{fa} \otimes (h_{fa} - h_{fx}))\]  
\[= (-R_{xx} \otimes R_{yy}) (h_{fa} \otimes h_{fx})\]  
(A13)
As $r_y(x)$ is even, (8) gives $r_{x'}(x) = -r_{x'}(-x)$ and $R_{x'x'} = -R_{x'x'}$. Therefore, using (A2)

$$R_{x'x'}h_{x'} \otimes R_y(h_{x'} - h_{y'f_y}) = R_{x'x'}h_{x'} \otimes R_yh_{y'f_y}$$

(A14)

which is satisfied for

$$R_y(h_{x'} - h_{y'f_y}) = R_yh_{y'f_y}$$

(A15)

which is true because

$$R_yh_{x'} = r_y = R_yh_{y'f_y} + R_{y'y}h_{y'f_y}$$

(A16)

where the right equality is by the symmetric version of (A5) and (A6) (replacing $x$ by $y$).

By symmetry (replacing $x$ with $y$), the equations involving $r_{yf_y}$ on the right-hand side of (A10) are also satisfied.

REFERENCES


Søren Forchhammer was born in 1959 in Copenhagen, Denmark. He received the M.S. degree in engineering and the Ph.D degree from the Technical University of Denmark, Lyngby, in 1984 and 1988, respectively. Currently, he is an Associate Professor at the Institute of Circuit Theory and Telecommunication at the Technical University of Denmark, where he has been employed since 1988. His main interests include data compression, image communications, and halftone images.

Kim S. Jensen was born in 1963 in Middelfart, Denmark. He received the M.S. degree in electrical engineering from the Technical University of Denmark, Lyngby, in 1989. From 1989 to 1992, he worked at the Institute of Circuit Theory and Telecommunication at the Technical University of Denmark, working with image processing for graphical applications.