ARCHITECTURES FOR GENERALIZED 2D FIR FILTERING USING SEPARABLE FILTER STRUCTURES

M. S. Andrews

LOGIC Devices, Inc.
1320 Orleans Ave.
Sunnyvale, CA 94089

ABSTRACT

The problem of generalized 2D FIR filtering for large filter kernel sizes can be computationally prohibitive when required in real-time, such as with video applications. In this paper, we describe architectures and design methods for generalized 2D FIR filtering employing LOGIC Devices' LF33xx (HV Filter) family of devices. The LF33xx family of devices is designed to perform dimensionally separate filtering (row/column) along orthogonal axes (horizontal direction and vertical direction for video applications). Additionally, we will briefly review 2D filter design and kernel separability theory.

1. INTRODUCTION

It is possible to achieve arbitrarily complex 2D filtering through the use of separability theory [1]. In this paper, we describe design techniques and practical architectures to achieve filtering with arbitrary two dimensional filtering kernels.

We show it is possible to design an arbitrary two dimensional (2D) filter and then separate the kernel dimensionally into a sum of orthogonal one dimensional (1D) filter coefficient sets. This is accomplished with the use of either an eigenvalue expansion of the 2D kernel or with the use of the Singular Value Decomposition (preferred method) [2].

In this note, we demonstrate the use of the Singular Value Decomposition (SVD) in separating 2D kernels into sums of 1D outer products so that the sum can be implemented on parallel or serial 1D filter configurations. Following relevant mathematical discussion, we proceed with an example kernel design and then to architectural implementations.

2. 2D FIR KERNEL DESIGN - A BRIEF REVIEW

Myriad techniques exist for the design of 2D FIR filter kernels. Consult [1] for more details. Two popular techniques are the window method and the frequency transformation (or McClellan transformation) method [1]. There are also various proposed optimal least squares type methods [3]. The techniques range from fairly simple to quite complex. It is only necessary to discuss one specific 2D FIR filter kernel design technique in order to illustrate the principles here. The window method is chosen here to illustrate the technique.

The window method is derived from the 1D case [1]. That is, a desired frequency response, $H_d(\omega_1, \omega_2)$, is specified in the ideal case and then a smooth window (e.g. Hamming, Blackman-Harris, etc.) [4] is applied to the 2D ideal frequency response

$$H(\omega_1,\omega_2) = H_d(\omega_1,\omega_2) * W(\omega_1,\omega_2) \qquad (1)$$

where * is the convolution operator and $W(\omega_1, \omega_2)$ is the frequency response of the chosen window function. The 2D filter kernel must of course be respecified in the spatial domain if spatial domain filters are to be used. This leads to a spatial specification as

$$h(n_1,n_2) = h_d(n_1,n_2)w(n_1,n_2) \qquad (2)$$

where $h(n_1, n_2)$ is the desired 2D filter kernel, $h_d(n_1, n_2)$ is the spatial function of the desired filter response and $w(n_1, n_2)$ is the spatial specification of the window. The 2D window function is easily computed from a 1D window function as an outer product, $w(n_1)w(n_2)$, where w is identical in n_1 and n_2. A comprehensive window table can be found in [4] and [5].

3. KERNEL SEPARATION VIA ORTHOGONAL DECOMPOSITION

It is fairly well known that an arbitrary 2D function can be decomposed into a sum of outer products of 1D orthogonal functions [2]. This is represented as

$$h(n_1,n_2) = \sum_{k=1}^{P} h_{k1}(n_1)h_{k2}(n_2) \qquad (3)$$

where $h(n_1, n_2)$ is the 2D kernel function and $h_{k1}(n_1)$ and $h_{k2}(n_2)$ are orthogonal 1D kernel sequences.

The 2D filter kernel is separated most generally with the use of the Singular Value Decomposition (SVD) [6]. The SVD of an arbitrary real matrix, $\mathbf{X}_{m\,x\,n}$, is compactly represented as

$$\mathbf{X}_{m\times n} = \mathbf{U}_{m\times n}\Sigma_{m\times n}\mathbf{V}_{n\times n}^{\ T} \qquad (4)$$

where $\mathbf{U}_{mxm}$ and $\mathbf{V}_{nxn}$ are orthonormal such that

$$\mathbf{U}_{m\times m}\mathbf{U}_{m\times m}^T = \mathbf{U}_{m\times m}^T\mathbf{U}_{m\times m} = \mathbf{I}_{m\times m} \qquad (5)$$

and

$$\mathbf{V}_{n\times n}\mathbf{V}_{n\times n}^T = \mathbf{V}_{n\times n}^T\mathbf{V}_{n\times n} = \mathbf{I}_{n\times n} \qquad (6)$$

and $\mathbf{\Sigma}_{mxn}$ is a diagonal matrix of singular values whos rank (which is equivalent to the number of non-zero diagonal elements) determines P in the sum of (3). A shorthand notation is often used for $\mathbf{\Sigma}$ so that whenever $m < n$,

$$\Sigma = diag(\sigma_1, \sigma_2, \cdots \sigma_m) \quad (7)$$

The separation of the 2D kernel, **H**, into the 1D kernel sequences, h_{k1} and h_{k2} (k = 1, 2, … P) is accomplished by the extraction of m columns of **U** and **V** once the SVD has been computed on the 2D filter kernel, **H**, according to (4). This suggests the following design steps:

1. Choose a 2D filter kernel design method and design the kernel, **H**
2. Apply the SVD to **H** according to (4) so that $\mathbf{H} = \mathbf{U\Sigma V}^T$
3. Determine the 2D kernel rank (P = the number of non-zero σ_i)
4. Extract P columns from **U** into h_{k1} sequences for the horizontal/row filters
5. Extract P columns from **V** into h_{k2} sequences for the vertical/column filters

First it should be noted that since **U** and **V** are orthonormal, the 1D sequences are already orthonormalized. The square root of the energy has been removed into σ_I. It should also be noted that various approximations to the filter **H** can be achieved by taking $P' < P$. This can provide a very useful approximation since a very large subclass of 2D filters fit into P = 2 space. P = 2 space is sufficient to represent most two-fold and four-fold symmetry constraints.

3.1 Filter Coefficient Design Example

In this section, we use MATLAB® [7] to perform the various calculations necessary to completely specify a 2D filter in terms of 1D coefficient sets. In keeping with the steps listed above, a 2D filter kernel is first designed using a 2D filter design technique.

Step 1: Design a 2D FIR Filter Using the 2D Window Method

We use the following design parameters for the basic 2D filter design:

1. Order = 15
2. Cutoff = 0.3 (Nyquist = 1)
3. Window = Hamming

The result of the filter design with the above parameters is plotted in Figure 1.

It may not be apparent from the examination of the mesh in Figure 1 what the rank of **H** might be. Recall from the discussion above that the rank of the matrix will exactly determine P. Occassionally the rank of **H** can be predicted based on careful examination of the contour of **H**. The contour

of **H** is shown in Figure 2 below where it can be seen that four-fold symmetry is present. When this type of symmetry condition exists a second order separation can almost surely be predicted.

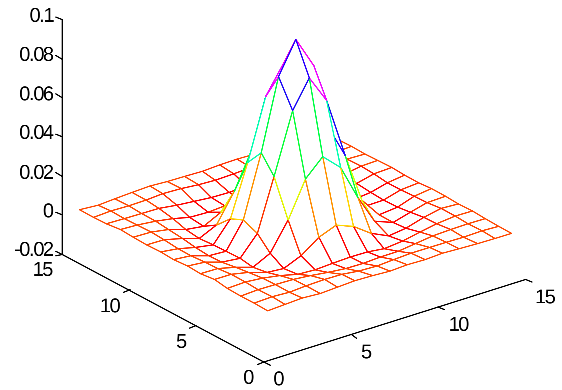

Figure 1: 2D FIR Filter Kernel Designed by 2D (Hamming) Window Method

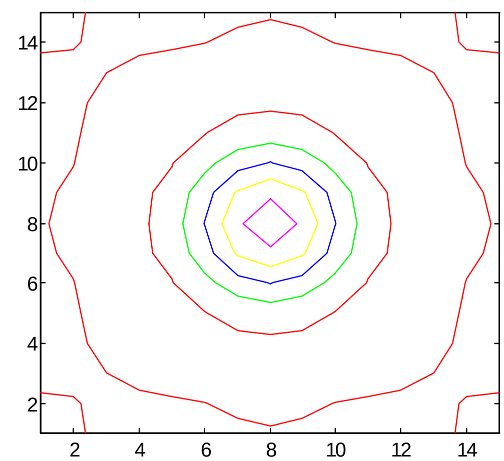

Figure 2: Contour of 2D FIR Filter Kernel H

Step 2: Compute the SVD of H Using (4)

Step 3: Determine the Rank of H from Σ

The diagonal of $\mathbf{\Sigma}$ has only two non-zero elements (0.2546 and 0.0220) in this case. This also determines the rank of **H** as two.

Step 4: Extract P = 2 Columns from U

The coeficients for the horizontal filter are contained in the first two columns of **U**, however, somewhere in the coefficient sets, the appropriate gain terms (0.2546 and 0.0220) must be applied. Either **U** or **V** (or both when square-rooted) can contain these terms. If they are applied to **U**, the resulting sets are graphed in Figure 3.

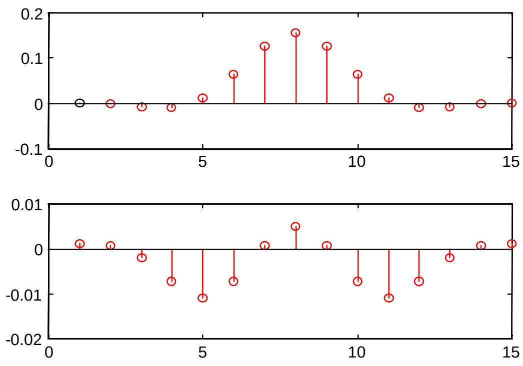

Figure 3: Normalized Horizontal Coefficients

Of course, these coefficients should be appropiately quantized for the target DSP device.

Step 5: Extract *P* = 2 Columns from V

The identical procedure is performed on **V** for the vertical coefficients. In this case, we quantize to 12 bit 2's complement format and show the results in Figure 4.

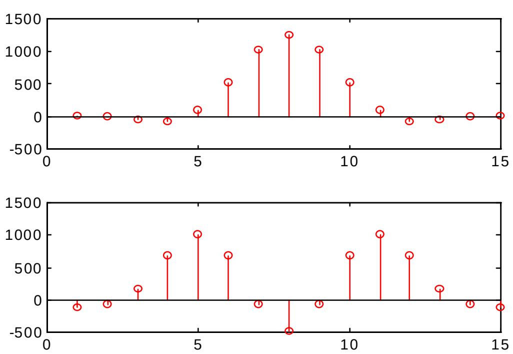

Figure 4: 12-bit Quantized Vertical Coefficients

This completes the coefficient design problem. It is now of interest to examine the architectural configurations necessary for performing the 2D filter operation specified by the kernel of Figure 1 and the coefficient sets of Figures 3 and 4.

4. HARDWARE CONFIGURATIONS

In this section, various architectures for performing generalized 2D convolution filtering employing the LOGIC Devices' LF3310, LF3320, and LF3330, along with their tradeoffs, are explored. The various architectures are the usual serial cascade, parallel cascade, interleaved, and single device interleaved.

The basic building block for multi-chip architectures capable of handling 15 x 15 kernel sizes consists of one LF3320 device and two LF3330 devices and one external 12 bit line buffer. This is shown in Figure 6.

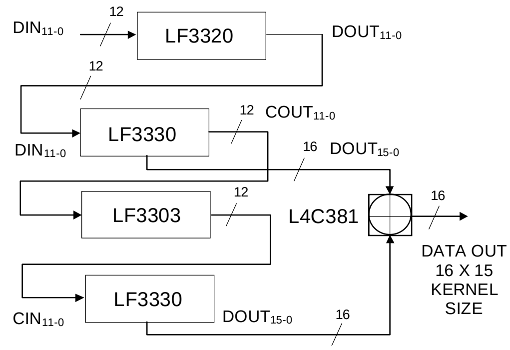

Figure 5: HV Building Block for 16 x 15 Separable Kernels

Whenever any of the coefficient sets is assymmetric (as can be the case in certain 2D separable kernels - especially for even kernel sizes e.g. 16 x 16), the full hardware of Figure 5 must be deployed. For symmetric coefficient sets, certain efficiencies are possible with the HV family of devices. For the general case, however, assymmetry is assumed. Also, it is important to understand that in the general assymmetric case, coefficients should be loaded in reverse manner (relative to data flow) to properly affect the convolution being performed.

If bus simplicity is a concern and the full 80 MHz rate these devices are capable of is required, the 15 x 15 kernel of Figure 1 can be implemented in two serial cascades of the basic HV building blocks as illustrated in Figure 6. Numerical concerns can arise when serial cascades are used and so it is important that the full dynamic range of the output of the first stage be available to the second stage via proper use of the RSL (Round, Select, and Limit) circuit present on all LF33xx devices (and so on if more stages are needed). The reader is referred to the data sheets of individual devices [8] on how to properly set up this subfunction. The only potential drawback of the system of Figure 6 is that additional latency will occur at the final output. This could be an issue for some systems.

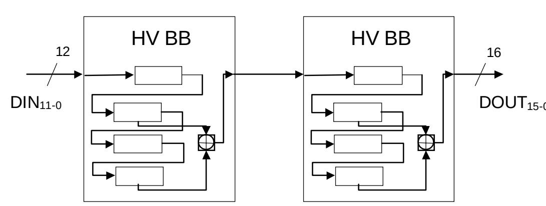

Figure 6: Serial Cascade for 2nd Order Separable 16 x 15 Kernels

For systems that demand the least latency in the throughput and the greatest possible speed, the parallel cascade can be used. This is illustrated in Figure 7. The only special requirement is for proper data alignment into the L4C381 ALU device.

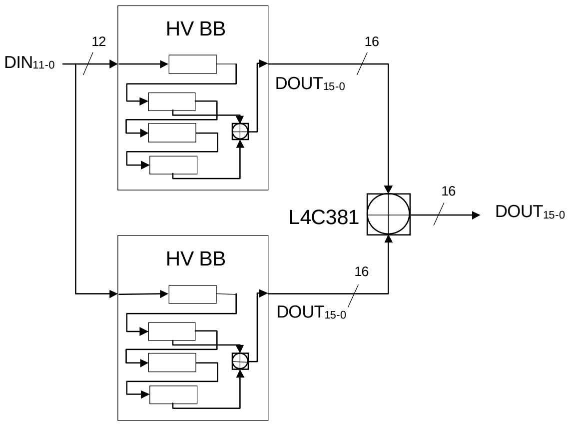

Figure 7: Parallel Cascade for 2nd Order Separable 16 x 15 Kernels

Since the LF33xx family of HV filters has the flexibility for interleaving data streams, it may be possible to take advantage of only one building block to implement second order separable kernels up to 16 x 15. This is possible because up to 16 fully independent streams can run through the block (of which only two streams are needed for second order separable kernels). The LF3330 can handle interleaved streams with the one caveat that the maximum pixels per line will be cut in half for each doubling of streams. Also, the rate will obviously be half of full rate for the two stream case. This implies that 40 MHz is the maximum rate for second order separable kernel filtering streams run interleaved. The extra hardware required for interleaving so that second order separable kernel filtering can be achieved is displayed in Figure 8.

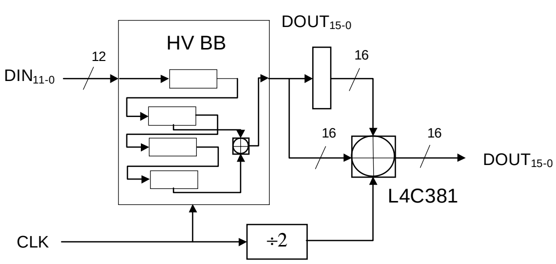

Figure 8: Interleave Operation for 2nd Order Separable 16 x 15 Kernels

Based on the discussion in the last paragraph, smaller assymmetric kernel sizes (maximum of 16 x 8) can be implemented with a single LF3310 in interleaved mode. Again, for second order separable kernels, the maximum data rate is 40 MHz. The same external hardware configuration as that shown in Figure 8 applies for this case.

Note it is possible to design coefficient sets for even ordered kernels also. This produces even symmetric coefficient sets as opposed to odd symmetric coefficient sets as has been designed here.

5. ACCURACY OF METHOD AND CONCLUDING REMARKS

The accuracy of the separation method for 2D convolution kernels depends on the order (or degree) of separability and the chosen order with which to represent it, as discussed before. In full floating point, accuracy is very close to machine precision. Figure 9 displays the FP error surface between the two filtered versions (full 2D and separable 2D) of a popular MRI test image.

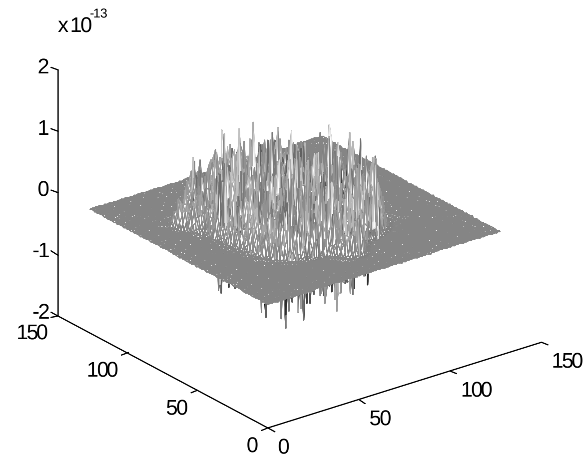

Figure 9: Error Surface Between 2D FIR and HV FIR for a popular MRI Test Image

REFERENCES

[1] Lim, J. S., *Two-Dimensional Signal and Image Processing*, Prentice-Hall, Englewood Cliffs, NJ 1990.

[2] Lu, W. S., H. P. Wang, and A. Antoniou, "Design of Two-Dimensional FIR Digital Filters by Using the Singular Value Decomposition," *IEEE Trans. Circuits Syst.*, Vol. 37, p. 35, Jan. 1990.

[3] Lodge, J. H. and M. F. Fahmy, "An efficient lp Optimization Technique for the Design of Two-Dimensional Linear Phase FIR Digital Filters," *IEEE Trans. Acoust., Speech and Sig. Proc.*, Vol. ASSP-28, pp. 308-313, June 1980.

[4] Harris, F. J., "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proc. IEEE,* vol. 66, pp. 51-83, Januray 1978.

[5] Proakis, J. G. and D. G. Manolakis, *Digital Signal Processing: Principles, Aglorithms, and Applications*, MacMillan Publishing Company, New York 1992.

[6] Klema, V. C. and A. J. Laub, "The Singular Value Decomposition: Its Computation and Some Applications," *IEEE Trans. Autom. Control*, Vol. AC-25, pp. 164-176, April 1980.

[7] Thompson, C. M. and Shure, L., *Matlab® Image Processing Toolbox User's Guide*, The Mathworks, Inc., Natick, MA 1995.

[8] http://www.logicdevices.com/