NEAREST-NEIGHBOR IMAGE MODEL

Aliaksei Sandryhaila, José M. F. Moura

Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

ABSTRACT

We propose a novel, adaptive model for image representation. The model places image pixels on the nodes of a two-dimensional nearest-neighbor graph. Edge weights for the graph depend on the image of interest, and can be determined by solving a corresponding least-squares problem. The proposed model is shown to provide an efficient image representation well-suited for image compression.

***Index Terms*—** Image representation, orthogonal transform, compression, nearest-neighbor graph, orthogonal polynomials.

1. INTRODUCTION

Images are ubiquitous examples of 2-D discrete signals that are used and studied in multiple disciplines. Effective image representations have been developed for various purposes, including coding, storage and transmission, recognition and classification, restoration and enhancement [1]. These techniques pursue different objectives, and they can be optimized to satisfy different performance requirements.

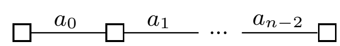
Some widely-used image representations are based on expanding images into orthonormal bases with the expectation that most information is captured with few basis functions. The expansion coefficients are calculated using a corresponding orthogonal transform. The optimal basis and transform for the representation and compression of a set of images is given by the corresponding Karhunen-Loève transform (KLT) [2]. However, there is no general efficient algorithm to compute this transform. In practice, images are often represented using other, more computationally-efficient orthogonal transforms, such as the discrete cosine (DCT) and discrete wavelet (DWT) transforms. These transforms are very efficient in the representation of natural, "smooth" images, and have given rise to widely-used JPEG and JPEG 2000 image compression standards [1].

Contributions. In this paper, we propose a novel representation of images that is based on two-dimensional (2-D) nearest-neighbor (NN) graphs. This work is based on the signal processing framework developed for signals represented with one-dimensional (1-D) NN graphs [3, 4]. We represent image pixels as nodes of a 2-D regular grid, which we view as 2-D undirected weighted NN graphs. An example of such a graph is shown in Fig. 1(b). A 2-D NN graph can be constructed as a Cartesian product of two 1-D NN graphs shown in Fig. 1(a). The weights of the graph edges are specific to images of interest and are constructed in an adaptive fashion. We introduce the corresponding signal representation model for signals on 2-D NN graphs and derive the corresponding orthonormal basis for image representation. As a potential application, we consider image compression and demonstrate the advantages of the proposed representation when compared to the widely-used DCT and DWT.

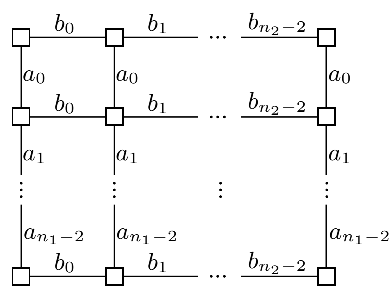
2. BACKGROUND

In this section, we discuss the signal model based on a 1-D NN graph. This model and the corresponding signal processing framework have been introduced in [3, 4]. In general, a 1-D NN graph is a weighted, directed line graph with possible self-links. However, in this paper, we only consider a simpler case of an undirected NN graph with no self-links, such as the graph shown in Fig. 1(a). Here, we review relevant properties and structures of the corresponding algebraic signal model and processing framework.

This work was supported in part by ONR grant MURI-N000140710747.



(a)



(b)

Fig. 1. Undirected, weighted NN graphs. a) 1-D graph; b) 2-D graph obtained as a Cartesian product of two 1-D graphs.

Signal model. Consider a finite discrete signal

$$\mathbf{s} = \begin{pmatrix} s_0 & s_1 & \dots & s_{n-1} \end{pmatrix}^T. \quad (1)$$

The signal processing framework for the analysis of such signals is determined by the corresponding *algebraic signal model* given as a triple $(\mathcal{A}, \mathcal{M}, \Phi)$. Here, $\mathcal{A}$ is a set of filters that is closed under the operations of serial and parallel connections, as well as amplification of filters. $\mathcal{M}$ is a set of signals that is closed under the operations of linear combination of signals and filtering by filters from $\mathcal{A}$. Finally, Φ is a generalized form of z-transform that maps discrete signals $\mathbf{s} \in \mathbb{C}^n$ in (1) to signals $s \in \mathcal{M}$.

As demonstrated in [3, 4], the signal model for 1-D discrete signals of length n residing on an 1-D weighted undirected NN graph 1(a) with edge weights[1] $a_0, \dots, a_{n-2}$ and no self-links is

$$\begin{aligned} &\mathcal{A} = \mathcal{M} = \mathbb{C}[x]/P_n(x), \\ &\Phi: \mathbb{C}^n \to \mathcal{M}, \ \mathbf{s} \mapsto s(x) = \textstyle\sum_{0 \le k < n} s_k P_k(x). \end{aligned} \quad (2)$$

It is called a *1-D finite discrete-NN model*. Here, polynomials $P_k(x)$ in (2) are *orthogonal polynomials* [5] that satisfy the recursion

$$x \cdot P_k(x) = a_{k-1} P_{k-1}(x) + a_k P_{k+1}(x), \quad (3)$$

[1]Weights a_k may be given or may need to be adapted according to an application. Coefficient selection in this work is discussed in Section 3.

with $P_{-1}(x) = 0$ and $P_0(x) = 1$. Each $P_n(x)$ is a polynomial of degree n with distinct real roots $\alpha_0, \dots, \alpha_{n-1}$. Both the signal and filter space in (2) are given by the polynomial algebra $\mathbb{C}[x]/P_n(x)$, which is a set of polynomials of degree less than $n = \deg P_n(x)$ with polynomial multiplication performed modulo $P_n(x)$. Hence, filters and signals are represented with polynomials $h(x) \in \mathcal{A}$ and $s(x) \in \mathcal{M}$. The operations of parallel filter connection and signal addition are defined as polynomial addition. The operations of serial filter connection and filtering of a signal with a filter are defined as polynomial multiplication modulo $P_n(x)$.

Fourier transform. The Fourier transform in the 1-D finite NN model (2) is calculated as

$$S_m = \sum_{k=0}^{n-1} P_k(\alpha_m) s_k. \tag{4}$$

The matrix form of (4) is $\mathbf{S} = \begin{pmatrix} S_0 & S_1 & \dots & S_{n-1} \end{pmatrix}^T = \mathbf{P}_{p,\alpha}\mathbf{s}$, where $\mathbf{s}$ is given by (1) and the matrix

$$\mathbf{P}_{P,\alpha} = \begin{pmatrix} P_0(\alpha_0) & \cdots & P_{n-1}(\alpha_0) \\ \vdots & \vdots & \vdots \\ P_0(\alpha_{n-1}) & \cdots & P_{n-1}(\alpha_{n-1}) \end{pmatrix} \tag{5}$$

is called a *discrete NN transform*. As shown in [4],

$$\mathbf{P}_{P,\alpha}\mathbf{P}_{P,\alpha}^T = a_{n-1} \cdot \operatorname{diag}\left(P_{n-1}(\alpha_k)P_n'(\alpha_k)\right)_{0 \le k < n} = \mathbf{D}_P \,.$$

Hence, the scaled discrete NN transform

$$\mathbf{D}_P^{-1/2}\,\mathbf{P}_{P,\alpha} \tag{6}$$

is an orthogonal matrix. There exist efficient computational algorithms for discrete NN transforms [6].

Shift matrix. Each filter $h(x) \in \mathcal{A}$ is a polynomial in x, and the filtering of a signal $s(x) \in \mathcal{M}$ by a filter $h(x) \in \mathcal{A}$ in the NN model (2) is represented as polynomial multiplication modulo $P_n(x)$. Hence, the basic non-trivial filtering operation in (2) is given by $x \cdot s(x) \mod P_n(x)$. We call this operation the *shift*.

Consider the matrix $\phi(x)$, such that the (i,j)th element of $\phi(x)$ equals a_i if $|i-j| = 1$ and zero otherwise:

$$\phi(x) = \begin{pmatrix} & a_0 & & \\ a_0 & & \ddots & \\ & \ddots & & a_{n-2} \\ & & a_{n-2} & \end{pmatrix}. \tag{7}$$

The shift operation and the matrix (7) are related as

$$\begin{aligned} u(x) &= x \cdot s(x) \mod P_n(x) \\ &= \Phi\big(\phi(x)\mathbf{s}\big). \end{aligned}$$

That is, the coefficients u_k of the resulting signal $u(x) = \sum_{k=0}^{n-1} u_k P_k(x)$ can be calculated as

$$\mathbf{u} = \begin{pmatrix} u_0 & u_1 & \dots & u_{n-1} \end{pmatrix}^T = \phi(x)\mathbf{s},$$

where $\mathbf{s}$ and $\phi(x)$ are given by (1) and (7). The matrix $\phi(x)$ in (7) is called the *shift matrix*.

3. IMAGE REPRESENTATION WITH NEAREST-NEIGHBOR GRAPHS

In this section, we introduce a novel signal model for images that is based on 2-D NN graphs. The model places image coefficients (pixel values) on a 2-D NN graph, such as the one shown in Fig. 1(b). We view this graph as the Cartesian product of two 1-D NN graphs shown in Fig. 1(a). The weights of the graph are determined in an adaptive way from the image pixel values.

Signal model. Signal models for the representation of multidimensional signals can be constructed as tensor products of 1-D models [4, 7]. Since $n_1 \times n_2$ images can be viewed as 2-D finite discrete signals, we construct the corresponding 2-D finite NN model as a tensor product of two 1-D finite NN models (2) for $n = n_1$ and $n = n_2$ as follows. Consider a $n_1 \times n_2$ signal

$$\mathbf{s} = \begin{pmatrix} s_{0,0} & s_{0,1} & \dots & s_{0,n_2-1} \\ \vdots & \vdots & \vdots & \vdots \\ s_{n_1-1,0} & s_{n_1-1,1} & \dots & s_{n_1-1,n_2-1} \end{pmatrix} \in \mathbb{C}^{n_1 \times n_2}. \tag{8}$$

The corresponding 2-D finite NN model is given by

$$\begin{aligned} &\mathcal{A} = \mathcal{M} = \mathbb{C}[x,y]/\langle P_{n_1}(x), Q_{n_2}(y)\rangle \\ &\Phi:\ \mathbb{C}^{n_1 \times n_2} \to \mathcal{M} \\ &\qquad \mathbf{s} \mapsto s(x,y) = \sum_{k_1=0}^{n_1-1}\sum_{k_2=0}^{n_2-1} s_{k_1,k_2} P_{k_1}(x) Q_{k_2}(y). \end{aligned} \tag{9}$$

Here, $P_{k_1}(x)$ and $Q_{k_2}(y)$ are orthogonal polynomials generated by recursions (3) with coefficients $a_0, \dots, a_{n_1-2}$ and $b_0, \dots, b_{n_2-2}$, respectively. In this model, filters and signals are two-variate polynomials $h(x,y)$ and $s(x,y)$. Multiplication is performed modulo the ideal $\langle P_{n_1}(x), Q_{n_2}(y)\rangle$ [4]. Since the model (9) is a tensor product of two 1-D models, this is equivalent to performing multiplication modulo $P_{n_1}(x)$ and $Q_{n_2}(y)$ simultaneously.

Let $\alpha_0, \dots, \alpha_{n_1-1}$ be the roots of $P_{n_1}(x)$; and $\beta_0, \dots, \beta_{n_2-1}$ be the roots of $Q_{n_2}(y)$. The Fourier transform

$$\mathbf{S} = \begin{pmatrix} S_{0,0} & S_{0,1} & \dots & S_{0,n_2-1} \\ \vdots & \vdots & \vdots & \vdots \\ S_{n_1-1,0} & S_{n_1-1,1} & \dots & S_{n_1-1,n_2-1} \end{pmatrix}$$

for the 2-D model (9) is calculated as

$$\mathbf{S} = \mathbf{P}_{P,\alpha}\mathbf{s}\mathbf{P}_{Q,\beta}^T,$$

where $\mathbf{P}_{P,\alpha}$ and $\mathbf{P}_{Q,\beta}$ are discrete NN transforms (5) for polynomials $P_{k_1}(x)$ and $Q_{k_2}(y)$, respectively.

The model has two basic shift operations that correspond to multiplication by x and y. In particular, signal coefficients u_{k_1,k_2} obtained as a result of the combined shift

$$u(x,y) = (x+y) \cdot s(x,y) \mod \langle P_{n_1}(x), Q_{n_2}(y)\rangle$$

can be calculated as

$$\mathbf{u} = \begin{pmatrix} u_{0,0} & \dots & u_{0,n_2-1} \\ \vdots & \vdots & \vdots \\ u_{n_1-1,0} & \dots & u_{n_1-1,n_2-1} \end{pmatrix} = \phi(x)\mathbf{s} + \mathbf{s}\phi(y)^T. \tag{10}$$

Shift matrices $\phi(x)$ and $\phi(y)$ are $n_1 \times n_1$ and $n_2 \times n_2$ matrices (7) with coefficients $a_0, \dots, a_{n_1-2}$ and $b_0, \dots, b_{n_2-2}$, respectively.

Coefficient selection. In order to construct a specific instantiation of the 2-D NN model (9) for an image of interest, we need to select appropriate coefficients $a_0, \ldots, a_{n_1-2}$ and $b_0, \ldots, b_{n_2-2}$. We determine these coefficients by constructing the shift matrices $\phi(x)$ and $\phi(y)$ that minimize the ℓ_2 distortion of the shift (10)[2]:

$$\left\{ \begin{array}{lll} a_0, & \ldots, & a_{n_1-2}, \\ b_0, & \ldots, & b_{n_2-2} \end{array} \right\} = \underset{a_{k_1}, b_{k_2} \in \mathbb{C}}{\operatorname{argmin}} ||\phi(x)\mathbf{s} + \mathbf{s}\phi(y)^T - \mathbf{s}||_2.$$

It can be solved as an overdetermined least-squares problem

$$\begin{pmatrix} A_0 & B_0 \\ \vdots & \vdots \\ A_{n_2-1} & B_{n_2-1} \end{pmatrix} \begin{pmatrix} a_0 \\ \vdots \\ a_{n_1-2} \\ b_0 \\ \vdots \\ b_{n_2-2} \end{pmatrix} = \mathbf{s}_v. \tag{11}$$

Here, each A_{k_2} is a $(n_1) \times (n_1 - 1)$ matrix with $A_{k_2}(i, i) = s_{i+1,k_2}$, $A_{k_2}(i+1, i) = s_{i,k_2}$ for $0 \leq i < n_1 - 1$ and other elements equal to zero. Each B_{k_2} is a $(n_1) \times (n_2 - 1)$ matrix with $B_{k_2}(i, k_2 - 1) = s_{i,k_2-1}$, $B_{k_2}(i, k_2) = s_{i,k_2+1}$ for $0 \leq i < n_1$ and other elements equal to zero. The vector $\mathbf{s}_v \in \mathbb{C}^{n_1 n_2}$ is the column-first vectorization of $\mathbf{s}$, so that $\mathbf{s}_v(i + jn_1) = \mathbf{s}(i, j)$.

4. IMAGE COMPRESSION

The proposed model can potentially be used for compression, adaptive filtering, and denoising of images. In this paper we study image compression, which is an extensive research area in image processing with multiple approaches, algorithms, and standards [1].

Compression using orthonormal bases. Orthonormal bases have long been used for efficient image representation and compression. A suitable basis captures most of the image information with relatively few basis functions (that is, most of the image energy is stored in a few projection coefficients with large magnitudes). In addition, orthogonality of basis functions removes redundancy in the image representation.

In general, if vectors $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_{n-1}$ form an orthonormal basis in $\mathbb{C}^n$, a signal $\mathbf{s} \in \mathbb{C}^n$ in (1) can be compressed as follows. First, the coefficients of the projection of $\mathbf{s}$ on the basis are computed as

$$\mathbf{c} = \begin{pmatrix} c_0 & \cdots & c_{n-1} \end{pmatrix}^T = \mathbf{B}\,\mathbf{s}, \tag{12}$$

where the matrix

$$\mathbf{B} = \begin{pmatrix} \mathbf{b}_0 & \cdots & \mathbf{b}_{n-1} \end{pmatrix}^H \tag{13}$$

represents an orthogonal transform. Without loss of generality, we can assume the coefficients $c_0, \ldots, c_{n-1}$ are ordered by decreasing magnitude, so that $|c_0| \geq |c_1| \geq \ldots \geq |c_{n-1}|$. The compressed signal $\hat{\mathbf{s}}$ can be reconstructed as

$$\hat{\mathbf{s}} = \mathbf{B}^H\,\hat{\mathbf{c}}, \tag{14}$$

where the vector $\hat{\mathbf{c}} = \begin{pmatrix} c_0 & \cdots & c_{\ell-1} & 0 & \cdots & 0 \end{pmatrix}^T$ is obtained from $\mathbf{c}$ by keeping only coefficients $c_0, \ldots, c_{\ell-1}$. The corresponding compression ratio is $R(\epsilon) = n/\ell$.

For a fixed peak signal-to-noise ratio (PSNR) ρ with the corresponding mean squared error (MSE) ϵ, we achieve the highest compression ratio $R(\epsilon)$ by finding the smallest ℓ that satisfies [1]

$$\frac{1}{n}||s - \hat{s}||_2^2 \leq \epsilon. \tag{15}$$

For multiple signals $\mathbf{s}_0, \ldots, \mathbf{s}_{m-1} \in \mathbb{C}^n$ that require, respectively, $\ell_0, \ldots, \ell_{m-1}$ coefficients to satisfy (15) for the MSE ϵ, the *average compression ratio* $R_{avg}(\epsilon)$ is

$$R_{avg}(\epsilon) = \frac{nm}{\ell_0 + \ell_1 + \ldots + \ell_{m-1}}. \tag{16}$$

Identifying a suitable basis for an efficient representation and compression of images of interest is a non-trivial task. A number of such bases have been proposed. Some of the most successful and widely-used ones are the cosine basis and wavelet bases. These bases are well-suited for representation of natural, “smooth” images. The corresponding orthogonal transform matrices $\mathbf{B}$ in (13) that compute the projection coefficients are the DCT and the DWT.

Proposed algorithm. Since the scaled discrete NN transform (6) is an orthogonal matrix, its rows represent an orthonormal basis in $\mathbb{C}^n$. Hence, the rows of the tensor product[3]

$$\mathbf{D}_Q^{-1/2}\,\mathbf{P}_{Q,\beta} \otimes \mathbf{D}_P^{-1/2}\,\mathbf{P}_{P,\alpha} \tag{17}$$

of discrete NN transforms $\mathbf{P}_{P,\alpha}$ and $\mathbf{P}_{Q,\beta}$, introduced in Section 3, form an orthonormal basis in $\mathbb{C}^{n_1 n_2}$.

Based on this property, we propose the tensor product (17) of $\mathbf{P}_{P,\alpha}$ and $\mathbf{P}_{Q,\beta}$ as the compressing orthogonal transform $\mathbf{B}$ in (13). The proposed compression algorithm consists of four steps:

Step 1. Represent an image $\mathbf{s}$ in (8) in the column-first vectorized form $\mathbf{s}_v$, so that $\mathbf{s}_v(i + jn_1) = \mathbf{s}(i, j)$.

Step 2. Determine the coefficients $a_0, \ldots, a_{n_1-2}$ and $b_0, \ldots, b_{n_2-2}$ by solving the least-squares problem (11).

Step 3. Construct the corresponding scaled discrete NN transforms $\mathbf{D}_P^{-1/2}\,\mathbf{P}_{P,\alpha}$ and $\mathbf{D}_Q^{-1/2}\,\mathbf{P}_{Q,\beta}$. Compute the projections $\mathbf{c}$ in (12) of $\mathbf{s}_v$ on the orthonormal basis given by the rows of the tensor product (17).

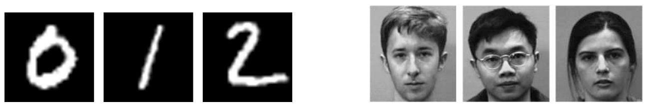
Step 4. For a fixed MSE ϵ, determine and keep the minimal required number of projection coefficients, discard others, and calculate the compressed signal $\hat{\mathbf{s}}$ in (14).

Compression of multiple images. One advantage of transforms such as the DCT and the DWT is that they do not depend on the image that we wish to compress. The proposed transform (17), however, depends on the image of interest. When compressing multiple images of the same size and similar structure and intensity, we can avoid constructing separate transforms for each image by considering a single 2-D NN model (9) for all images. The corresponding weights $a_0, \ldots, a_{n_1-2}$ and $b_0, \ldots, b_{n_2-2}$ are obtained by solving the following joint minimization problem for all m images $\mathbf{s}_0, \ldots, \mathbf{s}_{m-1}$ simultaneously:

$$\left\{ \begin{array}{lll} a_0, & \ldots, & a_{n_1-2}, \\ b_0, & \ldots, & b_{n_2-2} \end{array} \right\} = \underset{a_{k_1}, b_{k_2}}{\operatorname{argmin}} \sum_{i=0}^{m-1} ||\phi(x)\mathbf{s}_i + \mathbf{s}_i\phi(y)^T - \mathbf{s}_i||_2^2. \tag{18}$$

[2]This coefficient selection is partially motivated by [8], where images are modeled as non-causal autoregressive fields with constant coefficients determined by the MMSE prediction error. They can be visualized as 2-D NN graphs in Fig. 1(b) with edge weights $a_0 = \ldots = a_{n_1-2}$ and $b_0 = \ldots = b_{n_2-2}$. Our approach can extend the model in [8] to fields with non-constant coefficients. Image-based optimization of edge weights a_i and b_i also distinguishes our approach from other works, such as [9, 10], where edge weights are determined by vertices rather than signal coefficients assigned to vertices.

[3]A tensor product of matrices $A \in \mathbb{C}^{n_1 \times n_1}$ and $B \in \mathbb{C}^{n_2 \times n_2}$ is a matrix $C = A \otimes B \in \mathbb{C}^{n_1 n_2 \times n_1 n_2}$ such that $C(k_1 n_1 + k_2, m_1 n_1 + m_2) = A(k_1, m_1)B(k_2, m_2)$.



(a) Digits (b) Faces

Fig. 2. Examples of images used for the evaluation of the proposed compression algorithm.

5. EXPERIMENTS

Setup. To analyze the performance of the proposed algorithm, we apply it to the compression of two classes of images: hand-written digits and faces (examples are shown in Fig. 2). The digit images are obtained from the MNIST dataset [11]. We use 1000 images for each of the ten digits. The face images are obtained from the Faces94 dataset [12]. We use 20 images for each of 152 individuals.

For each set of images (1000 images per digit, 20 images per individual), we construct a 2-D NN model (9) by solving the joint minimization problem (18). We calculate the average compression ratio (16) for PSNR values $\rho \in \{30\text{ dB}, 40\text{ dB}, 50\text{ dB}, 60\text{ dB}\}$, which correspond to MSE values $\epsilon \in \{65, 6.5, 0.65, 0.065\}$ in (15). For comparison, we also consider two standard orthogonal transforms the DCT and the DWT, the latter based on orthogonal Daubechies filters[4] of length 4 with three decomposition levels [1]. The 2-D forms of both transforms, corresponding to the matrix $\mathbf{B}$ in (13), are constructed as tensor products of two 1-D transform matrices.

Discussion of results. Table 1 shows average compression ratios (16) obtained for different PSNR values. The proposed compression algorithm leads to increased average compression ratios in comparison to the DCT and the DWT.

Observe that both digit and face images are "smooth" and have relatively little sharp variations of the intensity in adjacent pixels. The DCT and the DWT are known to be well-suited for efficient representation and compression of such images. This observation makes the improvements in compression ratios achieved by the proposed algorithm more significant. Also, we use the method (18) for the construction of weights a_k and b_k. Other methods, such as a weighted ℓ_2-norm minimization or ℓ_0-norm minimization, may yield a more optimal choice of weights and lead to yet larger improvements in the representation efficiency and compression ratios.

6. CONCLUSIONS

We have proposed a novel, adaptive signal model for the representation of images. The model places image pixels on the nodes of a 2-D NN graph, obtained as the Cartesian product of two 1-D NN graphs. Edge weights for the graph depend on the image of interest, and can be constructed in an adaptive way by solving a corresponding least-squares problem. This representation uses and extends the theory of nearest-neighbor signal models. The proposed representation is shown to be well-suited for image compression and achieves higher compression ratios in comparison to other standard image compression techniques.

[4]Since we consider compression with orthonormal bases, we use an orthogonal DWT in our experiments. The JPEG 2000 standard uses biorthogonal wavelet bases that are beyond the scope of this paper.

Dataset	Algorithm	Peak signal-to-noise ratio			
		30 dB	40 dB	50 dB	60 dB
Digits	Proposed	5.8	3.0	2.4	2.1
	DCT	2.2	1.4	1.1	1.1
	DWT	5.2	3.0	2.3	1.9
Faces	Proposed	32.0	6.6	2.4	1.5
	DCT	4.4	1.8	1.3	1.1
	DWT	19.4	5.0	2.2	1.4

Table 1. Average compression ratios obtained for PSNR values of 30 dB, 40 dB, 50 dB, and 60 dB.

7. REFERENCES

[1] A. Bovik, *Handbook of Image and Video Processing*, Academic Press, 2nd edition, 2005.

[2] H. C. Andrews, "Multidimensional rotations in feature selection," *IEEE Trans. on Comp.*, vol. 20, no. 9, pp. 1045–1051, 1971.

[3] A. Sandryhaila, J. Kovacevic, and M. Püschel, "Algebraic signal processing theory: 1-D nearest-neighbor models," *IEEE Trans. on Signal Proc.*, vol. 60, no. 5, pp. 2247–2259, 2012.

[4] M. Püschel and J. M. F. Moura, "Algebraic signal processing theory," 2005, available at http://arxiv.org/abs/cs.IT/0612077, parts of this manuscript appeared as [13] and [14].

[5] W. Gautschi, *Orthogonal Polynomials: Computation and Approximation*, Oxford Univ. Press, 2004.

[6] J. R. Driscoll, D. M. Healy Jr., and D. Rockmore, "Fast discrete polynomial transforms with applications to data analysis for distance transitive graphs," *SIAM Journal Comput.*, vol. 26, pp. 1066–1099, 1997.

[7] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*, Prentice Hall, 1983.

[8] N. Balram and J. M. F. Moura, "Noncausal predictive image codec," *IEEE Trans. on Image Proc.*, vol. 5, no. 8, pp. 1229–1242, 1996.

[9] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *J. Appl. Comp. Harm. Anal.*, vol. 30, no. 2, pp. 129–150, 2011.

[10] S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter-banks for graph structured data," *IEEE Trans. on Signal Proc.*, vol. 60, no. 6, pp. 2786–2799, 2012.

[11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[12] L. Spacek, "Collection of facial images: Faces94," 2007, http://cswww.essex.ac.uk/mv/allfaces/faces94.html.

[13] M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: Foundation and 1-D time," *IEEE Trans. on Signal Proc.*, vol. 56, no. 8, pp. 3572–3585, 2008.

[14] M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: 1-D space," *IEEE Trans. on Signal Proc.*, vol. 56, no. 8, pp. 3586–3599, 2008.