

Data Representation using the Weyl Transform

Qiang Qiu*, Andrew Thompson†, Robert Calderbank*, and Guillermo Sapiro*

Abstract—The Weyl transform is introduced as a rich framework for data representation. Transform coefficients are connected to the Walsh-Hadamard transform of multiscale autocorrelations, and different forms of dyadic periodicity in a signal are shown to appear as different features in its Weyl coefficients. The Weyl transform has a high degree of symmetry with respect to a large group of multiscale transformations, which allows compact yet discriminative representations to be obtained by pooling coefficients. The effectiveness of the Weyl transform is demonstrated through the example of textured image classification.

Index Terms—Weyl transform, invariant representations, autocorrelation, Walsh-Hadamard transform, texture classification.

I. INTRODUCTION

Many signal processing tasks, such as detection, clustering, and classification, rely on representations which are invariant to a given group of transformations. Additional invariance to transformations which permute feature coefficients can often be achieved by pooling coefficients, for example Gabor wavelets [1] and scattering transforms [2].

Our focus here is not on new bases for representing signals, rather it is on the type of measurement that is fundamental to widely used algorithms for signal processing tasks. Thus, our focus is autocorrelation, and we show that autocorrelations can be calculated from trace inner products of covariance matrices with signed permutation matrices from the discrete Heisenberg-Weyl group. When a signal is transformed by an element of the (much larger) discrete symplectic group, these autocorrelations are fixed up to permutation and sign change. The symplectic group can be viewed as a discrete approximation to the full unitary group, and because it is so large (the order is $2^{m^2} \cdot (2^{2m} - 1)(2^{2m-2} - 1) \dots (2^2 - 1)$ for a signal of length 2^m), it allows for great versatility in the design of pooling strategies. Our approach makes use of the power of autocorrelation for describing multiscale periodicity in signals, and we demonstrate through the example of texture classification that this framework is useful for signal representation.

The mapping between a signal and its autocorrelation coefficients based on the Heisenberg-Weyl group is called the *Weyl transform*. This instance of the Weyl transform is a special case of a general framework for representation of operators in

harmonic analysis. In radar the larger (continuous) framework is fundamental to the study of the radar ambiguity function [3], [4], [5]. The binary Heisenberg-Weyl group studied here plays an important role in coding theory [6]. In this paper we make a new connection to signal analysis by describing the Weyl transform in terms of the Walsh-Hadamard transform [7] of binary autocorrelations.

Autocorrelation is known to be a powerful tool for detecting periodicity while imposing shift invariance, and has been used in a variety of signal processing tasks, including but not restricted to speech coding [8], pitch estimation [9] and noise removal [10]. Applications in image processing include character recognition, face detection, texture classification, and pattern recognition [11], [12], [13], [14]. Autocorrelation can be cyclic, or dyadic [15], the latter being especially well suited to representing multiscale texture. The Walsh-Hadamard Transform (WHT) is also well suited to periodic signals since it captures binary harmonics of data, being the binary counterpart of the Discrete Fourier Transform (DFT). The WHT can be combined with the Wiener-Khinchine convolution theorem to provide a fast method for calculating dyadic autocorrelation [16], and the windowed WHT has been used for pattern matching of images [17]. While all of the aforementioned previous work involving autocorrelation and the WHT is to some extent heuristic, we develop for the first time a rigorous mathematical theory of invariance which combines the two notions.

Fig. 1 illustrates how effectively the Weyl transform is able to encode differences in texture. It displays five randomly sampled 16×16 patches from each of two different textures. Each patch is broken down into smaller 4×4 sub-patches, their Weyl transforms calculated, and their absolute values aggregated over all sub-patches. Fig. 1(g) displays a similarity matrix where darker entries indicate greater Euclidean distance between Weyl transforms of different patches. Adjacent patches form clusters that clearly separate the two textures.

Section II introduces the Weyl transform and its invariance property, namely that the absolute values of Weyl coefficients are invariant to certain symmetries of the signal (Theorem II.1). We also establish a connection between the Weyl transform and the WHT of dyadic autocorrelations (Theorem II.3). In Section III, we illustrate the theory with examples of real-world textures. We demonstrate the versatility of the Weyl transform by pooling coefficients to obtain features with additional invariance to 90° rotation and cyclic translations. We also describe a supervised learning example in which training data is used to select the Weyl coefficients that are most significant in distinguishing the classes. For both examples, we compare our approach with three state-of-the-art image representations: Gabor wavelets [1], HOG [18] and LBP [19]. On the examples tested, the Weyl transform

The work of Calderbank and Thompson is supported in part by AFOSR under grant FA 9550-13-1-0076 and by NGA under grant HM017713-1-0006. Shapiro and Qiu are supported by NSF, ONR, AFOSR, NGA, and ARO.

*Q. Qiu, R. Calderbank and G. Sapiro are with the Department of Electrical Engineering, Duke University, USA; email: {qiang.qiu, robert.calderbank, guillermo.sapiro}@duke.edu;

†A. Thompson is with the Mathematical Institute, University of Oxford, UK; email: thompson@maths.ox.ac.uk

Q. Qiu and A. Thompson have contributed equally to this work.

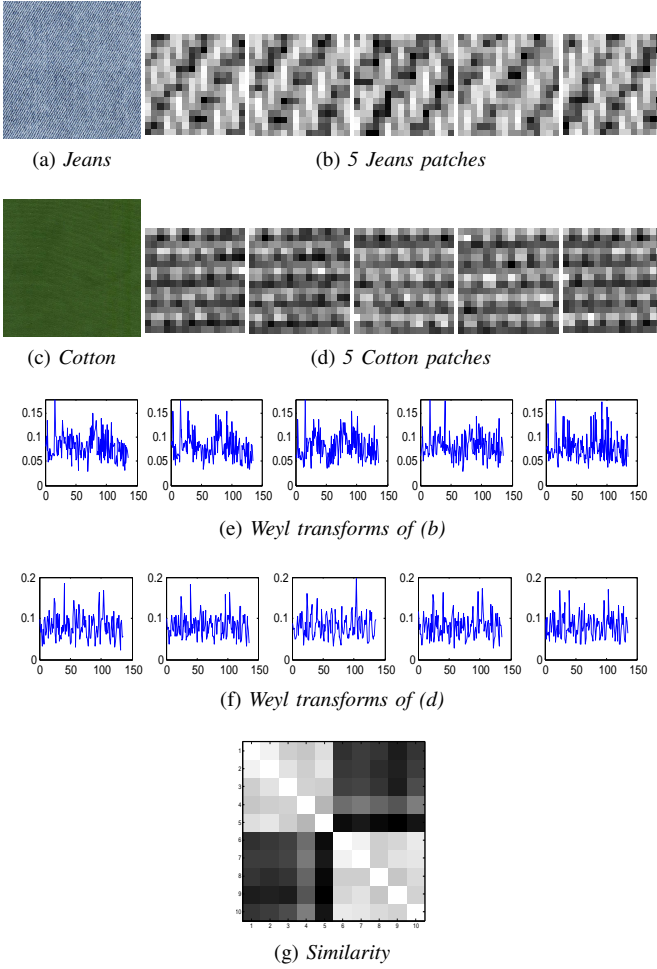


Fig. 1: The Weyl transform examples. (b) and (d) are five 16×16 patches randomly sampled from two textures (a) and (c), respectively. (e) and (f) are the Weyl transforms of (b) and (d); see Section II for details. (g) shows the similarity matrix for the Weyl transforms of patches from *Jeans* followed by *Cotton*, where darker colors indicate lower similarity. As shown in (g), patches sampled from the same texture share similar Weyl transforms, even though they exhibit obvious dissimilarity, e.g., due to translation.

regularly outperforms these other methods while using a significantly shorter feature vector. Full proofs of all results can be found in Section IV.

II. WEYL TRANSFORM THEORY

We now introduce some new fundamental theory about the Weyl transform. We first give a summary of the theory in Section II-A, in which we define the Weyl transform and give a crucial result about its invariance to certain multiscale transformations. The Weyl transform consists of inner products with matrices from the binary Heisenberg-Weyl group, and we describe this group of matrices more fully in Section II-B. In Section II-C, we make a connection between the Weyl transform and the Walsh-Hadamard Transform. Note that [6], [3], [4] show material related to the discrete Weyl transform, but presented in a different context, less suited to the data and applications described here.

A. The discrete Weyl transform

Given a vectorized signal $y \in \mathbb{R}^{2^m}$, we define its *Weyl coefficients* $\omega_{a,b}(y)$ to be

$$\omega_{a,b}(y) := \frac{1}{2^{m/2}} \text{Tr}[yy^T \cdot D(a,b)],$$

where $a = (a_{m-1} \dots a_0)^T$ and $b = (b_{m-1} \dots b_0)^T$ are binary m -tuples, and where the $\{D(a,b)\}$ are multiscale signed permutation matrices from the binary Heisenberg-Weyl group (see Section II-B for more details). We will denote the set of binary m -tuples by \mathbb{Z}_2^m . Figure 2 gives examples of matrices $D(a,b)$ for $m = 4$.



Fig. 2: Examples of the matrices $D(a,b)$ for $m = 4$: -1 (black), 0 (grey), 1 (white).

The mapping from covariance matrices $yy^T \in \mathbb{R}^{2^m \times 2^m}$ to vectors of Weyl coefficients in $\mathbb{R}^{2^{2m}}$ is an isometry, which we will refer to as the *Weyl transform*. A main result of this paper is that, whenever a $D(a,b)$ matrix is applied to a signal vector y , the magnitudes of its Weyl coefficients are unchanged. The Weyl transform is therefore invariant to a large class of multiscale signed permutations, a desirable property for a representation designed to detect periodicity.

Theorem II.1 (Weyl transform invariance property). *Let $\{\omega_{a,b}(y)\}$ be the Weyl coefficients of $y \in \mathbb{R}^{2^m}$. If $y' = D(a',b')y$ for some $(a',b') \in \mathbb{Z}_2^m$, then, for all $(a,b) \in \mathbb{Z}_2^m$ such that $a^T b = 0^1$, $|\omega_{a,b}(y')| = |\omega_{a,b}(y)|$.*

Proof: See Section IV-A.

As an illustration of the above result, Figure 3 displays a simple texture pattern, along with four patches taken from it at different positions and under different orientations. In the context of texture classification, it is desirable to have a representation which identifies all of these patches as coming from the same texture. In fact, all four patches can be obtained from each other by applying $D(a,b)$ matrices, and they therefore have the same Weyl transform coefficients in absolute value.

In Section I, we presented an example demonstrating the ability of the Weyl transform to effectively classify patches from two different textures (Fig. 1). The Weyl transform distinguishes the two textures by capturing multiscale symmetries which are important in describing texture. Moreover, invariance to multiscale transformations (Theorem II.1) ensures that the Weyl transforms of patches from the same texture exhibit similarity.

Each $D(a,b)$ matrix is a product of a permutation matrix and a diagonal ‘sign-change’ matrix with ± 1 entries. The sign-change patterns turn out to be columns of the Walsh-Hadamard transform matrix, also known as Walsh functions [7]. We will show in Section II-C that the Weyl transform can equally be viewed as a combination of autocorrelations and the Walsh-Hadamard transform, two familiar tools in signal processing (see Section I). This connection also provides a method for

¹The result in fact holds for all (a,b) , but we make the assumption $a^T b = 0$ to simplify some of the analysis.

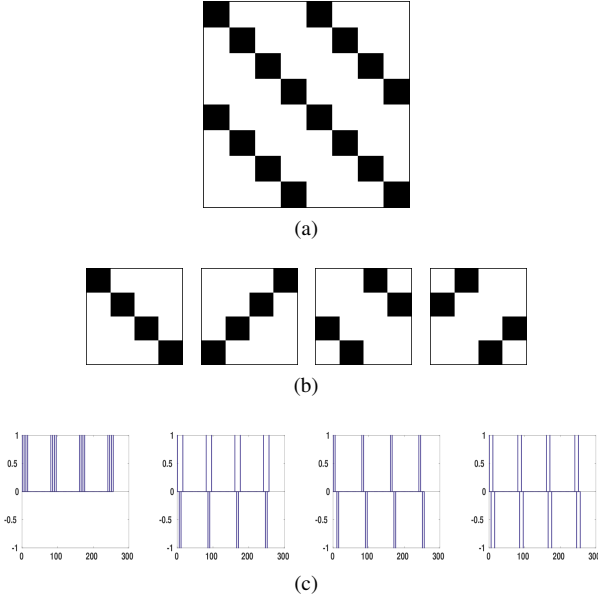


Fig. 3: (a) A simple texture pattern; (b) Four subpatches taken at different positions and orientations; (c) The Weyl transforms of each subpatch.

the fast computation of the Weyl transform by means of the fast (fully binary) Walsh-Hadamard transform (FWHT).

B. The binary Heisenberg-Weyl group

The Weyl transform consists of inner products with signed permutation matrices from the binary Heisenberg-Weyl group, which we next describe in more detail. The Heisenberg-Weyl group can be thought of as a multiscale extension of the Dihedral group D_8 , namely the symmetries of the square: four rotations through multiples of 90° , and reflections in its four axes of symmetry. These transformations can equally be viewed as matrix multiplications. Setting

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (1)$$

the matrix group $\{\pm X^a Z^b : a, b \in \mathbb{Z}_2\}$ provides a representation of D_8 , where $\mathbb{Z}_2 := \{0, 1\}$.

Next we use binary m -tuples to label the entries of a vector of length 2^m for some positive integer m . The v -th coordinate is labeled by the binary expansion of v , and coordinates are ordered from right to left, so that $v = v_{m-1}2^{m-1} + \dots + v_0$ is represented by $(v_{m-1} \dots v_1 v_0)^T \in \mathbb{Z}_2^m$ (where, for example, $\mathbb{Z}_2^2 = \{(0 \ 0)^T, (0 \ 1)^T, (1 \ 0)^T, (1 \ 1)^T\}$). Given $a = (a_{m-1} \dots a_0)^T \in \mathbb{Z}_2^m$, define $D(a, 0)$ to be the permutation matrix given by the Kronecker product

$$D(a, 0) := (X^{a_{m-1}}) \otimes \dots \otimes (X^{a_0}).$$

Note that the $D(a, 0)$ are dyadic multiscale permutations, with the leftmost terms giving coarse-scale permutations and the rightmost terms giving fine-scale permutations. The first row of Fig. 4 displays the matrices $D(a, 0)$ in the case of $m = 4$. Similarly, given $b = (b_{m-1} \dots b_0)^T \in \mathbb{Z}_2^m$, define $D(0, b)$ to be the sign change matrix given by the Kronecker product

$$D(0, b) := (Z^{b_{m-1}}) \otimes \dots \otimes (Z^{b_0}).$$

The matrices $D(0, b)$ for $m = 4$ are displayed in the second row of Fig. 4. We will show in Section II-C that the sign patterns are the columns of the Walsh-Hadamard transform matrix, also known as Walsh functions [7]. Define $D(a, b)$, for $a, b \in \mathbb{Z}_2^m$, by

$$D(a, b) := D(a, 0)D(0, b), \quad (2)$$

to obtain a collection of signed permutation matrices. The third row of Fig. 4 gives some examples of matrices $D(a, b)$ for $m = 4$.

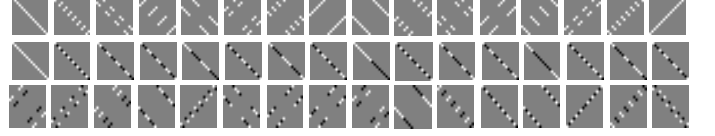


Fig. 4: The signed permutation matrices $D(a, b)$ for $m = 4$: -1 (black), 0 (grey), 1 (white). First row: Permutation matrices $D(a, 0)$. Second row: Sign change matrices $D(0, b)$. Third row: Examples of matrices $D(a, b)$.

Define

$$\mathcal{HW}_{2^m} := \{i^\lambda D(a, b) : a, b \in \mathbb{Z}_2^m, \lambda \in \mathbb{Z}_4\}. \quad (3)$$

The set \mathcal{HW}_{2^m} forms the binary Heisenberg-Weyl group under matrix multiplication [4].

The set of all real symmetric matrices is a vector space, which we denote by V . Given real symmetric matrices $R, S \in V$, associate with V the inner product $(R, S) := \text{Tr}(R^T S)$, which induces the Frobenius norm $\|R\|_F := [\text{Tr}(R^T R)]^{1/2}$ on V . We now show that matrices $D(a, b)$ with $a^T b = 0$ form a basis for V . Define

$$\mathbb{Y}_2^m := \{(a, b) : a, b \in \mathbb{Z}_2^m, a^T b = 0\},$$

and

$$\mathcal{B}_{2^m} := \left\{ \frac{1}{2^{m/2}} D(a, b) : a, b \in \mathbb{Y}_2^m \right\}.$$

Lemma II.2. \mathcal{B}_{2^m} is an orthonormal basis for V .

Proof: If $a^T b = 1$ then $D(a, b)$ squares to $-I$, and hence the inner product of $D(a, b)$ with any real symmetric matrix is zero. For more details see Section IV-A.

Lemma II.2 implies that any real symmetric matrix $R \in \mathbb{R}^{2^m \times 2^m}$ can be expanded as

$$R = \sum_{(a, b) \in \mathbb{Y}_2^m} \left\{ \frac{1}{2^{m/2}} \text{Tr}[R \cdot D(a, b)] \right\} \frac{1}{2^{m/2}} D(a, b).$$

In particular, given a vectorized signal $y \in \mathbb{R}^{2^m}$, its correlation matrix yy^T can be expanded as

$$\begin{aligned} yy^T &= \sum_{(a, b) \in \mathbb{Y}_2^m} \left\{ \frac{1}{2^{m/2}} \text{Tr}[yy^T \cdot D(a, b)] \right\} \frac{1}{2^{m/2}} D(a, b) \\ &= \sum_{(a, b) \in \mathbb{Y}_2^m} \omega_{a, b}(y) \frac{1}{2^{m/2}} D(a, b). \end{aligned} \quad (4)$$

The coefficients $\omega_{a, b}(y)$ are the *Weyl coefficients* of y as introduced in Section II-A. When we connect the Weyl transform to autocorrelation in Section II-C, it will be convenient to pad

with zeros by defining $\omega_{a,b} := 0$ when $a^T b = 1$. Since we have expanded in an orthonormal basis, the mapping from yy^T to $\{\omega_{a,b}\}$ is an isometry, namely the *Weyl transform* as introduced in Section II-A.

The Weyl transform also has a striking geometrical interpretation. Each $D(a,b)$ matrix (except for the identity) has ± 1 eigenspaces of multiplicity $m/2$ respectively, so that

$$\begin{aligned} D(a,b) &= [P_{a,b} \quad Q_{a,b}] \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \begin{bmatrix} P_{a,b}^T \\ Q_{a,b}^T \end{bmatrix} \\ &= P_{a,b} P_{a,b}^T - Q_{a,b} Q_{a,b}^T, \end{aligned}$$

where $P_{a,b}$ and $Q_{a,b}$ are orthonormal bases for the $+1$ and -1 eigenspaces respectively. The Weyl coefficient $\omega_{a,b}$ can therefore be expressed as

$$\begin{aligned} \omega_{a,b} &= \frac{1}{2^{m/2}} \text{Tr}[yy^T (P_{a,b} P_{a,b}^T - Q_{a,b} Q_{a,b}^T)] \\ &= \frac{1}{2^{m/2}} \left\{ \|P_{a,b}^T y\|_F^2 - \|Q_{a,b}^T y\|_F^2 \right\}, \end{aligned}$$

which reveals that $\omega_{a,b}$ gives information about the relative distance of the covariance matrix yy^T from two half-spaces. A large positive value means that the image is in the $+1$ eigenspace; a large negative value means that the image is in the -1 eigenspace. A large absolute value of either sign indicates that the image exhibits periodic symmetry.

It is shown in [20] that the collection of half-spaces induced by the Weyl transform is in fact the optimal half-space packing originally given in [21]². The Weyl transform can therefore be viewed as a principled matched filter for covariance matrices.

C. The Weyl transform and autocorrelation: a connection

We next establish the bridge between Weyl coefficients and autocorrelation, showing that the Weyl transform can equally be viewed in terms of the Walsh-Hadamard transform [7] of binary autocorrelations. Consider a vectorized signal $y \in \mathbb{R}^{2^m}$ indexed by a binary m -tuple $v = (v_{m-1} \dots v_0)^T$. Then the correlation matrix yy^T may be divided into 2^m autocorrelation bands $z_a \in \mathbb{R}^{2^m}$, where

$$(z_a)_v := y_v y_{v+a}. \quad (5)$$

Each autocorrelation band z_a gives information on the invariance of a signal to a particular binary translation $v \rightarrow v + a$. Given a binary m -tuple a , define $\omega_a \in \mathbb{R}^{2^m}$ to be the subset of Weyl coefficients indexed by a , that is $(\omega_a)_b := \omega_{a,b}$. Define the $2^m \times 2^m$ Walsh-Hadamard transform matrix H_{2^m} by

$$(H_{2^m})_{v,w} := \frac{1}{2^{m/2}} (-1)^{v^T w},$$

where v and w are binary m -tuples [7]. The following result characterizes the Weyl transform as a combination of autocorrelations and the Walsh-Hadamard transform.

Theorem II.3 (Weyl transform in terms of autocorrelation). *Let the autocorrelation bands of y , $\{z_a\}$, be defined as in (5). Then $\omega_a(y) = H_{2^m} z_a$, where $\{\omega_a\}_b = 0$ if $a^T b = 1$.*

Proof: See Section IV-B.

²It is also shown in [21] that this subspace packing can be constructed recursively.

III. ILLUSTRATIVE EXAMPLES: THE INVARIANCE AND VERSATILITY OF THE WEYL TRANSFORM

We next further illustrate fundamental properties of the Weyl transform using real-world texture examples. We suggest here two ways to exploit the Weyl transform for effective signal representation: equivalence class histograms and supervised coefficient selection. We first exploit the underlying group structure of the transform to build invariance to particular geometrically-significant transformations; and then we describe how training data can be used to select the Weyl coefficients that are most significant in distinguishing the classes.

A. Equivalence class histograms

While the absolute values of the Weyl coefficients are already invariant under Heisenberg-Weyl transformations, more can be said if we allow the coefficients to be permuted (but unchanged in absolute value). Given an arbitrary linear transformation $y \rightarrow \phi y$, the cyclic property of the trace implies that

$$\begin{aligned} \omega_{a,b}(\phi y) &= \frac{1}{2^{m/2}} \text{Tr}[\phi y y^T \phi^T D(a,b)] \\ &= \frac{1}{2^{m/2}} \text{Tr}[y y^T \phi^T D(a,b) \phi], \end{aligned}$$

which in turn implies that ϕ permutes the Weyl coefficient absolute values if it is an outer automorphism of \mathcal{HW}_{2^m} . The outer automorphism group of the Heisenberg-Weyl group is known: it is the binary symplectic group $\text{Sp}(2m, \mathbb{Z}_2)$ [4]. Based on these observations, we may conceive of a method of pooling Weyl coefficients which builds invariance to any desired symplectic transformation, which we next describe. While often, e.g., in deep learning, pooling is heuristically designed [22], here we provide a principled approach to it [23], [2].

Let G be some (sub)group of symplectic transformations. Then G acts on \mathcal{HW}_{2^m} by conjugation, and the action partitions \mathcal{HW}_{2^m} into equivalence classes. This means that the group G can only permute the elements of \mathcal{HW}_{2^m} within each equivalence class. It follows that the average of the absolute values of the Weyl coefficients within a given equivalence class is invariant under G . Such equivalence class averaging has the further appeal of reducing the number of Weyl coefficients, and it is also extremely versatile since any group of geometrically-significant symplectic transformations can be considered.

As an example, we illustrate the proposed approach in the context of a vectorized $N \times N$ image y , taking G to be the group of transformations generated by 90° rotation and cyclic horizontal and vertical translation by any multiple of $N/4$. We first show that these are each indeed symplectic transformations. We assume that our original image Y has dimensions $N \times N$, $N = 2^r$, and that it is vectorized columnwise to give y , a vector of length 2^m where $m = 2r$. Let $a, b \in \mathbb{Z}_2^{2r}$, and let us write $a = (a_{2r-1} \dots a_0)^T$ and $b = (b_{2r-1} \dots b_0)^T$. We write 1_r for the vector of r ones.

Proposition III.1 (Symplectic permutations). *(i) Rotation of Y by 90° clockwise corresponds to the mapping*

$$D\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right) \longrightarrow (-1)^{1^T b_1} D\left(\begin{bmatrix} a_2 \\ a_1 \end{bmatrix}, \begin{bmatrix} b_2 \\ b_1 \end{bmatrix}\right), \quad (6)$$



Fig. 5: Seven fabric textures.

where $a_1, a_2, b_1, b_2 \in \mathbb{Z}_2^r$.

(ii) *Cyclic translation of Y by $N/4$ vertically corresponds to the mapping*

$$D\left(\begin{bmatrix} a_1 \\ j \\ k \\ a_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ l \\ m \\ b_2 \end{bmatrix}\right) \longrightarrow (-1)^m D\left(\begin{bmatrix} a_1 \\ j+k \\ k \\ a_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ l \\ l+m \\ b_2 \end{bmatrix}\right), \quad (7)$$

where $a_1, b_1 \in \mathbb{Z}_2^r$, $a_2, b_2 \in \mathbb{Z}_2^{r-2}$ and $j, k, l, m \in \mathbb{Z}_2$.

Proof: See Section IV-C.

An analogous result for cyclic translation by $N/4$ horizontally can be obtained by permuting the vertical and horizontal coordinates. Note that a cyclic translation by $N/2$ fixes the Weyl coefficients' absolute values, revealing the transformation to be an element of the Heisenberg-Weyl group.

For illustration, in the following experiments, we consider the case of a 4×4 image patch, vectorized to give $y \in \mathbb{R}^{16}$, so that its Weyl coefficients $\omega_{a,b}$ are indexed by a pair of 4-tuples. Let G be the group of transformations generated by 90° rotation and cyclic translation by $N/4$ (in this case, all possible cyclic translations) either vertically or horizontally. G partitions the a coordinates into the 6 equivalence classes

$$\begin{aligned} &\{(0\ 0\ 0\ 0)\}, \{(1\ 0\ 1\ 0)\}, \{(0\ 0\ 1\ 0), (1\ 0\ 0\ 0)\}, \\ &\{(0\ 0\ 0\ 1), (0\ 0\ 1\ 1), (0\ 1\ 0\ 0), (1\ 1\ 0\ 0)\}, \\ &\{(0\ 1\ 0\ 1), (0\ 1\ 1\ 1), (1\ 1\ 0\ 1), (1\ 1\ 1\ 1)\}, \\ &\{(0\ 1\ 1\ 0), (1\ 0\ 0\ 1), (1\ 0\ 1\ 1), (1\ 1\ 1\ 0)\}, \end{aligned} \quad (8)$$

while, independently, G also partitions the b coordinates into the 6 equivalence classes

$$\begin{aligned} &\{(0\ 0\ 0\ 0)\}, \{(0\ 1\ 0\ 1)\}, \{(0\ 0\ 0\ 1), (0\ 1\ 0\ 0)\}, \\ &\{(0\ 0\ 1\ 0), (0\ 0\ 1\ 1), (1\ 0\ 0\ 0), (1\ 1\ 0\ 0)\}, \\ &\{(1\ 0\ 1\ 0), (1\ 0\ 1\ 1), (1\ 1\ 1\ 0), (1\ 1\ 1\ 1)\}, \\ &\{(1\ 0\ 0\ 1), (0\ 1\ 1\ 0), (0\ 1\ 1\ 1), (1\ 1\ 0\ 1)\}. \end{aligned} \quad (9)$$

We thus partition the Weyl coefficients into 36 equivalence classes.

Six of the classes contain Weyl coefficients which are all zero, since $a^T b = 1$ for all members of the class, and so we discard these classes. In addition, the six classes for which $a = 0$ give variance information rather than autocorrelation information, so we also discard these classes. We are left with 24 classes, within each of which we average the Weyl coefficient absolute values. Since G permutes the members of each class, these averages are invariant to all transformations in G . We have thereby reduced the number of Weyl coefficients from 136 to 24, creating a representation that is invariant under not only the Heisenberg-Weyl group, but also the group G . In particular, compactness means that different ways of orienting and translating the same texture produce the same descriptor. Note that this equivalent class histogram approach is versatile: any geometrically-significant symplectic transformations could

be quotiented out in this way depending on the expected image symmetries to be encountered.

Fabric texture examples. Seven real-world fabric textures are shown in Fig. 5. In this set of experiments, we randomly sample 500 16×16 -sized gray-scale patches from each texture, and mix them up. The obtained 3500 texture patches are represented using 5 different descriptors: Weyl, intensity (the gray-scale value), Gabor [1], LBP [19], and HOG [18]. With the equivalence classes defined above, 24 Weyl coefficients (absolute values) are required to represent each 4×4 patch. For a 16×16 patch, we further average across the 16 obtained 24-sized vectors to form a 24-sized Weyl descriptor. The Gabor features used here have 5 scales and 8 orientations, down-sampled by a factor of 4, and each feature vector is of length 640. We use the basic LBP histogram of length 256. For HOG features, we use a block size of 4×4 and 36 histogram bins, which gives a 576-sized feature vector. The proposed Weyl representation is therefore the most compact one.

In various texture analysis tasks, e.g., texture classification, segmentation, and retrieval, it is critical to have a consistent representation for patches of the same texture type; and it is also desirable to have for each patch a small size descriptor to enable efficient computation. Fig. 6 shows the obtained 3500 texture patches represented in different descriptors. Different textures are shown in different colors. For visualization, data are plotted with the dimension reduced to 2 using PCA; while the full-sized descriptors, without the PCA step, are used in the actual experiments. A good texture descriptor encourages a compact and isolated cluster for the same color points. To enable a quantitative assessment, we perform k-means clustering on texture patches represented in different descriptors (before dimension reduction is applied). The proposed Weyl representation is the most compact and discriminative: a 24-sized descriptor gives 86.49% clustering accuracy. Note that the clustering accuracy here only approximately assesses the discriminability of a representation, since different clustering schemes can lead to different clustering results. The popular HOG and LBP provide comparable discriminability, but require an order of magnitude longer descriptors than the Weyl representation. We notice that the Gabor descriptor fails to effectively represent patches from *Fabric* and *Textile Blue*. After excluding these two textures, we obtain 74.96% clustering accuracy using Gabor features still below our proposed representation.

We next obtain multiple variants of each texture in Fig. 5 using reflections in horizontal and vertical axes and rotations through multiples of 90° (dihedral transformations). We repeat the above experiments by sampling 500 patches from multiple

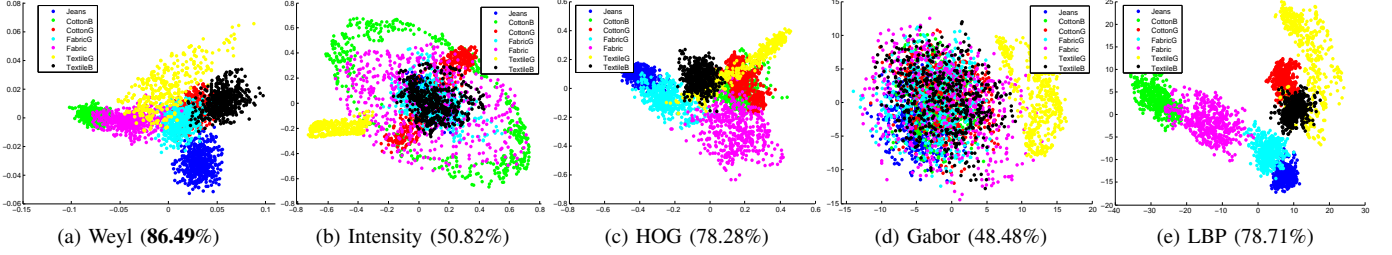


Fig. 6: Fabric texture patches represented with different descriptors. Data are visualized with the dimension reduced to 2 using PCA. Different textures are shown in different colors: *Jeans* (blue), *CottonB* (green), *CottonG* (red), *FabricG* (cyan), *Fabric* (magenta), *TextileG* (yellow), *TextileB* (black). The k-means clustering accuracies in parentheses approximately assess the discriminability (and class compactness) of each descriptor. The respective size of each descriptor is Weyl (24), Intensity (256), HOG (576), Gabor (640) and LBP (256). The proposed Weyl descriptor is both the most compact and discriminative.

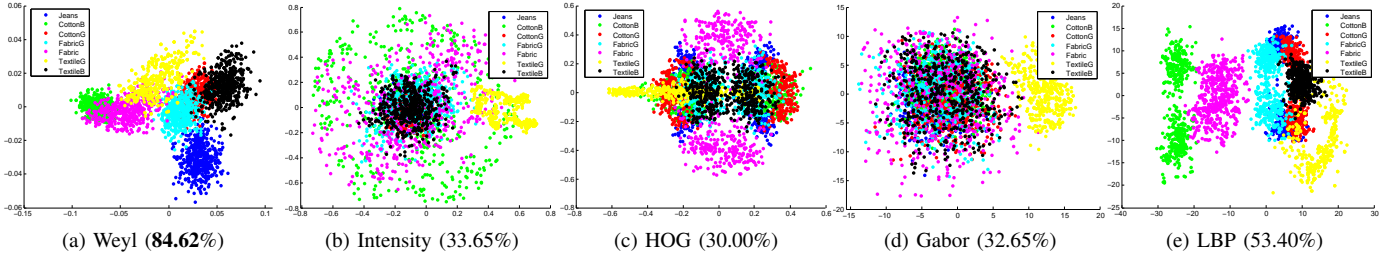


Fig. 7: Fabric texture patches under the dihedral transformations represented with different descriptors (see Fig. 6 for notation description). Comparing to Fig. 6, we observe that the Weyl representation is invariant under dihedral group transformations. For other descriptors, we observe new sub-clusters, indicating that significantly different descriptors are assigned to the same texture under dihedral transformations.

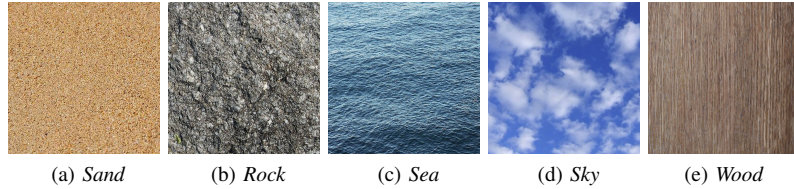


Fig. 8: Five nature textures.

variants of each texture. The sampled patches are represented using different descriptors, and plotted in Fig. 7 with different colors for different texture types. By comparing Fig. 7 to Fig. 6, we observe that the Weyl representation is invariant under dihedral transformations. For other descriptors, new sub-clusters emerge among the same color points; in other words, those descriptors assign significantly different representations to the same texture under different dihedral transformations, causing inferior performance for many texture analysis tasks.

Note that different feature extraction methods often focus on different information aspects. For example, among features compared in the paper, LBP captures texture, HOG captures shape, intensity captures color, and Gabor captures salient visual properties. Therefore, without a concrete application, there is often no clear winning feature in terms of discrimination. By comparing with several popular features, our main objective here is to illustrate properties of the Weyl representation, such as compactness, invariance and discrimination.

Textures from Nature. We conduct the same set of experiments on textures of significantly less regularity, i.e., the

five nature textures in Fig. 8. As shown in Fig. 9, the Weyl representation is still the most compact and discriminative. In Fig. 10, we again observe the invariance of the Weyl representation under dihedral transformations, enabling significantly more consistent representation than other descriptors.

B. Supervised coefficient selection

We suggest here another effective way to adopt the Weyl representation through supervised coefficient selection. Consider two-class data points (multi-class cases can be supported using the standard one-vs-all strategy), where the class labels are known beforehand for training purposes. Let Y_+ and Y_- denote the set of points in each of the two classes respectively, where points are arranged as columns of the corresponding matrix. We assume here that the same number of points are selected for each class for computational efficiency.

We now rank all Weyl coordinates based on the discriminability, i.e., how significant they are in distinguishing the classes, and only use the best K Weyl coordinates for classification. The ability of the Weyl coordinate (a, b) in separating

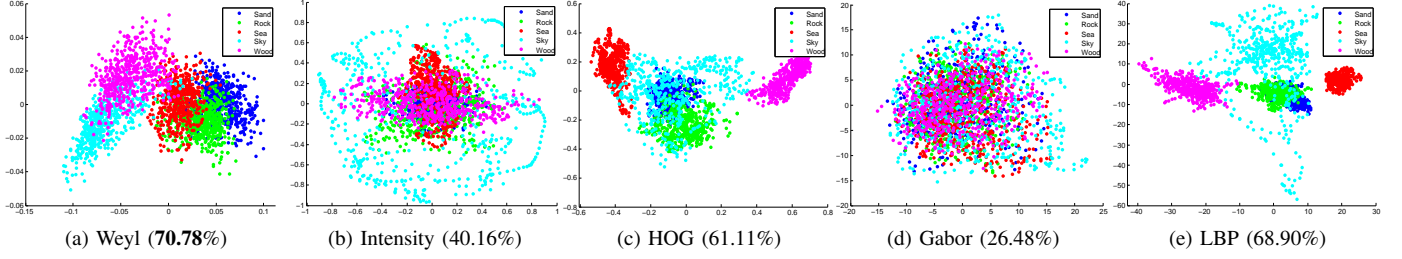


Fig. 9: Nature textures represented with different descriptors. For visualization, data are plotted with the dimension reduced to 2 using PCA. Different textures are shown in different colors: *Sand* (blue), *Rock* (green), *Sea* (red), *Sky* (cyan), *Wood* (magenta). The k-means clustering accuracies in the parentheses approximately assess the discriminability of each descriptor. The proposed Weyl descriptor is the most compact and discriminative.

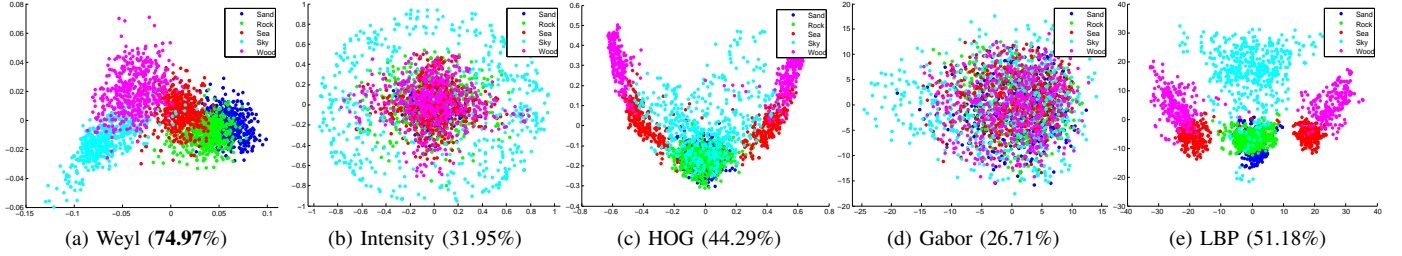


Fig. 10: Same as Fig. 7 for the textures in Fig. 8.

two classes can be assessed by the Weyl coefficient magnitude $|\omega_{a,b}(M)|$ of the matrix $M = (Y_+ - Y_-)(Y_+ + Y_-)'$: We pick one point from each class, $y_+ \in Y_+$ and $y_- \in Y_-$, with their respective Weyl coefficient difference at the coordinate (a, b) given by,

$$\begin{aligned}
 & |\text{Tr}[y_+ y_+^T D(a, b)] - \text{Tr}[y_- y_-^T D(a, b)]| \\
 &= |\text{Tr}[y_+^T D(a, b) y_+] - \text{Tr}[y_-^T D(a, b) y_-]| \\
 &= |y_+^T D(a, b) y_+ - y_-^T D(a, b) y_-| \\
 &= |(y_+ - y_-)^T D(a, b) (y_+ + y_-)| \\
 &= |\text{Tr}[(y_+ - y_-)(y_+ + y_-)^T D(a, b)]|. \quad (10)
 \end{aligned}$$

Thus, we can calculate the total Weyl coefficient difference between Y_+ and Y_- at the coordinate (a, b) as $|\text{Tr}[(Y_+ - Y_-)(Y_+ + Y_-)^T D(a, b)]|$. We rank all Weyl coordinates (a, b) in descending order of $|\omega_{a,b}(M)|$, and only use the best K coordinates for classification. Recalling the geometrical interpretation of Weyl coefficients in terms of relative distance from two half-spaces in Section II-B, the aim here is to identify the K subspace pairs whose relative distances discriminate the most between the two classes.

We pick a pair of texture images from Fig. 5 and 8, and randomly sample 500 16×16 -sized gray-scale patches from each texture. We randomly choose 20 patches for each texture as training, and the remaining 960 patches as testing. Patches are represented using different descriptors, and Nearest Neighbor classifiers are used to assign each testing patch to a texture type. Note that, for testing samples, only coefficients at the K selected Weyl coordinates need to be computed. Table 11 reports the results. The best Weyl coefficient already provides accuracies comparable to a 256 long intensity feature descriptor. When 16 Weyl coefficients are used, we obtain accuracies comparable to all other popular descriptors, but

with an order of magnitude shorter descriptor. Fig. 12 shows the classification accuracies when different numbers of best Weyl coefficients are used; note the fast convergence.

	size	Jeans vs. TextileB	Sand vs. Sea	Sand vs. Rock
Intensity	256	86.45	74.06	66.56
HOG	576	97.60	99.79	83.54
Gabor	640	91.14	71.25	64.79
LBP	256	100	100	85.41
Weyl	1	99.89	99.68	66.56
Weyl	3	100	99.68	76.25
Weyl	16	100	99.79	84.06

Fig. 11: Accuracy (%) of classifying texture patches represented using different descriptors.

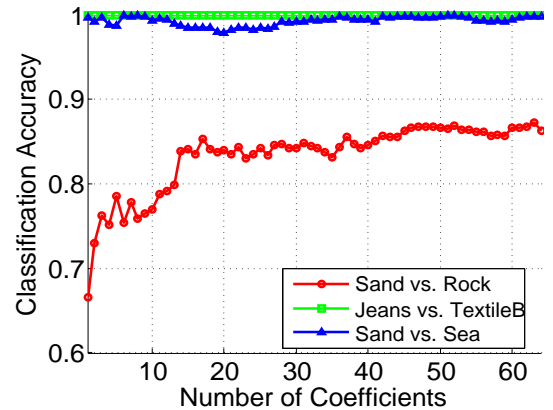


Fig. 12: Classification accuracies with different number of best Weyl coefficients.

IV. PROOFS

This section contains proofs of all results given in Sections II and III.

A. Proof of the Weyl transform invariance property

Recalling the definitions of the $D(a, b)$ matrices given in (2) and (1), we also introduce the notation

$$e_0 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad e_1 := \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

for the canonical basis elements in \mathbb{R}^2 . In order to prove Lemma 2.1 and Theorem 2.2 we need the following lemma, which gives various properties of the $D(a, b)$ matrices.

Lemma IV.1. *Let X , Z , and $D(a, b)$ be defined in (2) and (1). Then, for all $v, a, b, a', b' \in \mathbb{Z}_2^m$,*

$$(i) D(a, b)e_v = (-1)^{b^T v} e_{v+a}; \quad (11)$$

$$(ii) [D(a, b)]^T = [D(a, b)]^{-1} = (-1)^{a^T b} D(a, b); \quad (12)$$

$$(iii) D(a, b)D(a', b') = (-1)^{a'^T b} D(a + a', b + b'). \quad (13)$$

Proof: It follows from (1) that, given $v, a, b \in \mathbb{Z}_2$

$$(X^a Z^b)_v = (-1)^{bv} e_{v+a}. \quad (14)$$

Given $v = (v_{m-1} \dots v_0)^T \in \mathbb{Z}_2^m$, we have

$$e_{(v_{m-1} \dots v_0)} = e_{v_{m-1}} \otimes \dots \otimes e_{v_0}, \quad (15)$$

which combines with (2) and a standard property of the Kronecker product to give

$$\begin{aligned} D(a, b)e_v &= (X^{a_{m-1}} Z^{b_{m-1}}) \otimes \dots \otimes (X^{a_0} Z^{b_0}) (e_{v_{m-1}} \otimes \dots \otimes e_{v_0}) \\ &= (X^{a_{m-1}} Z^{b_{m-1}} e_{v_{m-1}}) \otimes \dots \otimes (X^{a_0} Z^{b_0} e_{v_0}). \end{aligned} \quad (16)$$

Combining (14) and (16) then gives

$$\begin{aligned} D(a, b)e_v &= [(-1)^{b_{m-1}v_{m-1}} e_{(v_{m-1}+a_{m-1})}] \otimes \dots \otimes [(-1)^{b_0v_0} e_{(v_0+a_0)}] \\ &= (-1)^{b^T v} (e_{(v_{m-1}+a_{m-1})} \otimes \dots \otimes e_{(v_0+a_0)}), \end{aligned}$$

and (11) now follows by (15). It follows from (11) that $[D(a, b)]^2 = I$ for all $a, b \in \mathbb{Z}_2^m$, from which it follows that $[D(a, b)]^T = [D(a, b)]^{-1}$. Inverting (11), and letting $w = v + a$, we obtain

$$[D(a, b)]^{-1} e_w = (-1)^{b^T(w+a)} e_{w+a} = (-1)^{a^T b} D(a, b)e_w$$

for all $w, a, b \in \mathbb{Z}_2^m$, which completes the proof of (12). We also have, by (11), that

$$D(a + a', b + b')e_v = (-1)^{(b+b')^T v} e_{v+a+a'} \quad (17)$$

for all $v, a, b, a', b' \in \mathbb{Z}_2^m$. But two applications of (11) give

$$\begin{aligned} D(a, b)D(a', b')e_v &= D(a, b)(-1)^{b'^T v} e_{v+a'} \\ &= (-1)^{b^T(v+a')} (-1)^{b'^T v} e_{v+a+a'} \\ &= (-1)^{a'^T b} (-1)^{(b+b')^T v} e_{v+a+a'}. \end{aligned} \quad (18)$$

Since (17) and (18) hold for all $v \in \mathbb{Z}_2^m$, (13) follows. \square

Proof of Lemma 2.1: A symmetric matrix is determined by the entries on and above the diagonal, which gives

$$\dim(V) = 2^m + \frac{1}{2} \cdot 2^m(2^m - 1) = \frac{1}{2} \cdot 2^m(2^m + 1). \quad (19)$$

On the other hand, writing $a = (a_{m-1} \dots a_0)^T, b = (b_{m-1} \dots b_0)^T$, conditioning on the pair (a_{m-1}, b_{m-1}) leads to the recurrence $|\mathbb{Y}_2^m| = 2|\mathbb{Y}_2^{m-1}| + 2^{2(m-1)}$, from which it is easy to deduce that $|\mathbb{Y}_2^m| = \frac{1}{2} \cdot 2^m(2^m + 1)$, which combines with (19) to give $|\mathbb{Y}_2^m| = \dim(V)$.

It remains to show that the elements of \mathcal{B}_{2^m} are orthonormal. It follows from (12) that, for any $(a, b) \in \mathbb{Y}_2^m$,

$$\text{Tr} \left\{ \frac{1}{2^{m/2}} [D(a, b)]^T \frac{1}{2^{m/2}} [D(a, b)] \right\} = \frac{1}{2^m} \text{Tr} [(-1)^{a^T b} I] = 1,$$

where in the last step we use the assumption that $a^T b = 0$. Similarly, it follows from (12) and (13) that, for any $a, b, a', b' \in \mathbb{Z}_2^m$,

$$\begin{aligned} &\text{Tr} \left\{ \frac{1}{2^{m/2}} [D(a, b)]^T \frac{1}{2^{m/2}} [D(a', b')] \right\} \\ &= \frac{1}{2^m} \text{Tr} [(-1)^{a^T b} (-1)^{a'^T b'} D(a + a', b + b')] \end{aligned} \quad (20)$$

If $a \neq a'$, $D(a + a', b + b')$ is zero on its diagonal, whereas if $a = 0$ and $b \neq 0$, the diagonal terms are nonzero but sum to zero. This observation combines with (20) to show that the elements of \mathcal{B}_{2^m} are orthogonal, which completes the proof. \square

Proof of Theorem 2.2 By (12) and (13), we have, for all $(a, b) \in \mathbb{Y}_2^m$,

$$\begin{aligned} &[D(a', b')]^T D(a, b) D(a', b') \\ &= (-1)^{a'^T b} [D(a', b')]^T D(a + a', b + b') \\ &= (-1)^{a'^T b'} (-1)^{a'^T b} D(a', b') D(a + a', b + b') \\ &= (-1)^{a'^T(b+b')} (-1)^{(a+a')^T b'} D(a, b) \\ &= (-1)^{a'^T b' + a'^T b} D(a, b), \end{aligned}$$

which proves the result. \square

B. Proof of Theorem 2.3

We may use (11) to deduce

$$D(a, b)y = \sum_v D(a, b)y_v e_v = \sum_v (-1)^{b^T v} y_v e_{v+a},$$

which gives

$$\{D(a, b)y\}_v = (-1)^{b^T(v+a)} y_{v+a}. \quad (21)$$

Meanwhile, the cyclic property of the trace gives

$$\text{Tr}[y y^T D(a, b)] = \text{Tr}[y^T D(a, b)y] = y^T D(a, b)y. \quad (22)$$

Combining (21) and (22), we have

$$\begin{aligned} \{\omega_a(y)\}_b &= \omega_{a,b} = \frac{1}{2^{m/2}} y^T D(a, b)y \\ &= \frac{1}{2^{m/2}} \sum_v y_v \{D(a, b)y\}_v \\ &= \frac{(-1)^{a^T b}}{2^{m/2}} \sum_v (-1)^{b^T v} y_v y_{v+a} \\ &= \frac{1}{2^{m/2}} \sum_w (-1)^{b^T w} y_w y_{w+a}, \end{aligned} \quad (23)$$

where in the last step we made the substitution $w = v + a$. Now (23) implies that $\{\omega_a(y)\}_b = 0$ whenever $a^T b = 1$, and the result now follows from the definitions of H_{2^m} and z_a . \square

C. Proof of Proposition 3.1

Rotation by 90° clockwise can be achieved by an interchange of vertical and horizontal coordinates followed by a reversal of the horizontal coordinates, which is the mapping

$$\theta : e_{(v_1, v_2)} \longrightarrow e_{(v_2+1_r, v_1)}, \quad (24)$$

where y is indexed by the binary $2r$ -tuple $v = (v_1 \ v_2)^T$, $v_1, v_2 \in \mathbb{Z}_2^r$, and e_v is a canonical basis vector. Writing $a = (a_1 \ a_2)^T$, $b = (b_1 \ b_2)^T$, we have by (24) and (11) that

$$\begin{aligned} & \theta^{-1} D \left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right) \theta e_{(v_1, v_2)} \\ &= \theta^{-1} D \left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right) e_{(v_2+1_r, v_1)} \\ &= (-1)^{1_r^T b_1 + b_1^T v_2 + b_2^T v_1} \theta^{-1} e_{(v_2+a_1+1_r, v_1+a_2)} \\ &= (-1)^{1_r^T b_1 + b_1^T v_2 + b_2^T v_1} e_{(v_1+a_2, v_2+a_1)}. \end{aligned} \quad (25)$$

Meanwhile, (11) also implies that

$$\begin{aligned} & (-1)^{1_r^T b_1} D \left(\begin{bmatrix} a_2 \\ a_1 \end{bmatrix}, \begin{bmatrix} b_2 \\ b_1 \end{bmatrix} \right) e_{v_1, v_2} \\ &= (-1)^{1_r^T b_1 + b_1^T v_2 + b_2^T v_1} e_{(v_1+a_2, v_2+a_1)}, \end{aligned}$$

which combines with (25) to give the result for rotation. Turning to the translation, consider the permutation

$$t : e_{(v_1, p, q, v_2)} \longrightarrow e_{(v_1, p+q, q+1, v_2)}, \quad (26)$$

where $v_1 \in \mathbb{Z}_2^r$, $v_2 \in \mathbb{Z}_2^{r-2}$ and $p, q \in \mathbb{Z}_2$. The mapping t permutes the first two vertical binary coordinates as

$$\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix},$$

which gives a vertical cyclic translation through $N/4$. Writing $a = (a_1 \ j \ k \ a_2)^T$, $b = (b_1 \ l \ m \ b_2)^T$, we have by (26) and (11) that

$$\begin{aligned} & t^{-1} D \left(\begin{bmatrix} a_1 \\ j \\ k \\ a_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ l \\ m \\ b_2 \end{bmatrix} \right) t e_{(v_1, p, q, v_2)} = \\ & t^{-1} D \left(\begin{bmatrix} a_1 \\ j \\ k \\ a_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ l \\ m \\ b_2 \end{bmatrix} \right) e_{(v_1, p+q, q+1, v_2)} = \\ & (-1)^{b_1^T v_1 + b_2^T v_2 + l(p+q) + m(q+1)} t^{-1} e_{(v_1+a_1, p+q+j, q+1+k, v_2+a_2)} \\ &= (-1)^{b_1^T v_1 + b_2^T v_2 + lp + lq + mq + m} e_{(v_1+a_1, p+j+k+1, q+k, v_2+a_2)}. \end{aligned} \quad (27)$$

Meanwhile, (11) also implies that

$$\begin{aligned} & (-1)^m D \left(\begin{bmatrix} a_1 \\ j+k \\ k \\ a_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ l+m \\ l \\ b_2 \end{bmatrix} \right) e_{v_1, p, q, v_2} \\ &= (-1)^{b_1^T v_1 + b_2^T v_2 + lp + lq + mq + m} e_{(v_1+a_1, p+j+k+1, q+k, v_2+a_2)}, \end{aligned}$$

which combines with (27) to give the result for translation. \square

V. CONCLUDING REMARKS

We have described how to represent measurement using the binary Weyl transform and developed new theory, supported by texture classification experiments, that connects mathematical properties of the transform with a broad class of signal processing objectives. In particular, we used invariance of the transform to a large class of multi-scale dihedral transformations to pool Weyl coefficients, thereby developing very concise histograms with high discriminative power. We also described a supervised method that learns the transform coefficient with the greatest discriminative value through training. Illustrative examples using real world texture examples were provided for both approaches.

We have demonstrated that the Weyl transform is a powerful tool for capturing image texture at the patch level. A promising direction for future work is to incorporate a patch-level Weyl transform into a deep learning framework, capable of learning high-level texture-based features at the image level.

REFERENCES

- [1] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. Academic Press, 2009.
- [2] L. Sifre and S. Mallat, "Rotation, scaling and deformation invariant scattering for texture discrimination," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.
- [3] S. Howard, A. Calderbank, and W. Moran, "Finite Heisenberg-Weyl groups and Golay complementary sequences," in *International Waveform Diversity and Design Conference*, 2006.
- [4] —, "The finite Heisenberg-Weyl groups in radar and communication," in *EURASIP Journal on Applied Signal Processing*, 2006, pp. 1–12.
- [5] W. Miller Jr, *Topics in harmonic analysis with applications to radar and sonar*, ser. IMA Volumes in Mathematics and its Applications. Springer-Verlag, 1991, pp. 66–168.
- [6] A. Calderbank, P. Cameron, W. Kantor, and J. Seidel, "Z4-kerdock codes, orthogonal spreads and Euclidean line-sets," *Proceedings of the London Mathematical Society*, vol. 75, no. 3, pp. 436–480, 1997.
- [7] S. Aghaian, H. Sarukhanyan, K. Egiazarian, and J. Astola, *Hadamard Transforms*. SPIE Press, 2011.
- [8] T. Ogunfunmi and M. Narasimha, *Principles of Speech Coding*. CRC Press, 2010.
- [9] P. McLeod and G. Wyvill, "A smarter way to find pitch," in *Proceedings of the International Computer Music Conference*, 2005.
- [10] H. Seck, Y. Zhang, and Y. Soh, "Autocorrelation noise removal for optical coherence tomography by sparse filter design," *Journal of Biomedical Optics*, vol. 17, no. 7, 2012.
- [11] D. Chetverikov and Z. Foldvari, "Affine-invariant texture classification using regularity features," in *Texture Analysis in Machine Vision*, ser. Machine Perception and Artificial Intelligence, M. Pietikainen, Ed. World Scientific, 2000.
- [12] R. Goudail, E. Lange, T. Iwamoto, K. Kyuma, and N. Otsu, "Face recognition system using local autocorrelations and multiscale integration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 1024–1028, 1996.
- [13] J. McLaughlin and J. Raviv, "Nth-order autocorrelations in pattern recognition," *Journal of Information Control*, vol. 12, pp. 121–142, 1968.
- [14] V. Popovici and J. Thiran, "Pattern recognition using higher-order local autocorrelation coefficients," *Pattern Recognition Letters*, vol. 25, no. 10, pp. 1107–1113, 2004.
- [15] J. Xiaoguang and M. Nixon, "Analysing front view face profiles for face recognition via the Walsh transform," *Pattern Recognition Letters*, vol. 15, no. 6, pp. 551–558, 1994.
- [16] R. Stankovic and M. Karpovsky, "Remarks on calculation of autocorrelation on finite dyadic groups by local transformations of decision diagrams," in *EUROCAST Conference on Computer Aided Systems Theory*, 2005, pp. 301–310.
- [17] Y. Hel-Or and H. Hel-Or, "Real-time pattern matching using projection kernels," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 9, pp. 1430–1435, 2005.

- [18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition*, June 2005.
- [19] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 657–662, 2006.
- [20] R. Calderbank, R. Hardin, E. Rains, P. Shor, and N. Sloane, "A group-theoretic framework for the construction of packings in grassmannian spaces," *Journal of Algebraic Combinatorics*, vol. 9, no. 2, pp. 129–140, 1999.
- [21] P. Shor and N. Sloane, "A family of optimal packings in grassmannian spaces," *Journal of Algebraic Combinatorics*, vol. 7, no. 2, pp. 157–163, 1998.
- [22] Y.-L. Boureau, N. Le Roux, F. Bach, J. Ponce, and Y. LeCun, "Ask the locals: Multi-way local pooling for image recognition," in *IEEE International Conference on Computer Vision*, 2011.
- [23] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.