Graph-based Transforms for Video Coding

Hilmi E. Egilmez, Member, IEEE, Yung-Hsuan Chao, Member, IEEE and Antonio Ortega, Fellow, IEEE

Abstract—In many state-of-the-art compression systems, signal transformation is an integral part of the encoding and decoding process, where transforms provide compact representations for the signals of interest. This paper introduces a class of transforms called graph-based transforms (GBTs) for video compression, and proposes two different techniques to design GBTs. In the first technique, we formulate an optimization problem to learn graphs from data and provide solutions for optimal separable and nonseparable GBT designs, called GL-GBTs. The optimality of the proposed GL-GBTs is also theoretically analyzed based on Gaussian-Markov random field (GMRF) models for intra and inter predicted block signals. The second technique develops edge-adaptive GBTs (EA-GBTs) in order to flexibly adapt transforms to block signals with image edges (discontinuities). The advantages of EA-GBTs are both theoretically and empirically demonstrated. Our experimental results demonstrate that the proposed transforms can significantly outperform the traditional Karhunen-Loeve transform (KLT).

Index Terms—Transform coding, predictive coding, graph-based transforms, video coding, compression, optimization, statistical modeling.

I. INTRODUCTION

Predictive transform coding is a fundamental compression technique adopted in many block-based image and video compression systems, where block signals are initially predicted from a set of available (already coded) reference pixels, and then the resulting residual block signals are transformed (generally by a linear transformation) to decorrelate residual pixel values for effective compression. After prediction and transformation steps, a typical image/video compression system applies quantization and entropy coding to convert transform coefficients into a stream of bits. Fig. 1 illustrates a representative encoder-decoder architecture comprising three basic components, (i) prediction, (ii) transformation, (iii) quantization and entropy coding, which are implemented in state-of-the-art compression standards such as JPEG [5], HEVC [6] and VP9 [7]. This paper focuses mainly on the transformation component of video coding and develops techniques to design orthogonal transforms, called graph-based transforms (GBTs), adapting diverse characteristics of video signals.

In predictive transform coding of video, the prediction is typically carried out by choosing one among multiple intra and inter prediction modes in order to exploit spatial and temporal redundancies between block signals. On the other hand, for the transformation, generally a single transform such as the discrete cosine transform (DCT-2) is applied in a separable manner to rows and columns of each residual block. The main problem of using fixed block transforms is the implicit assumption that all residual blocks share the same statistical properties. However, residual blocks can have very diverse statistical characteristics depending on the video content and the prediction mode (as will be demonstrated by some of the experiments in Section VII). A recent video coding standard, HEVC [6], partially addresses this problem by allowing the use of the asymmetric discrete sine transform (ADST or DST-7) in addition to the DCT-2 for small (i.e., 4 × 4) intra predicted blocks [8]. Yet, it has been shown that better compression can be achieved by using data-driven transform designs that adapt to statistical properties of residual blocks [9]–[19].

The majority of prior studies about transforms for video coding focus on developing transforms for intra predicted residuals. In [9], a mode-dependent transform (MDT) scheme is proposed by designing a Karhunen-Loeve transform (KLT) for each intra prediction mode. More recently in [10], [11], the MDT scheme is similarly implemented for the HEVC standard, where a single KLT is trained for each intra prediction mode offered in HEVC. Moreover in [12]–[17], the authors demonstrate considerable coding gains over the MDT method by using the rate-distortion optimized transformation (RDOT) scheme, which suggests designing multiple transforms for each prediction mode so that the encoder can select a transform (from the predetermined set of transforms) by minimizing a rate-distortion (RD) cost. Since the RDOT scheme allows the flexibility of selecting a transform on a per-block basis at the encoder side, it provides better adaptation to residual blocks with different statistical characteristics as compared to the MDT scheme. However, all of these methods rely on KLTs derived from sample covariance matrices, which may not be good estimators for the true covariances of models, especially when the number of data samples is small [20], [21]. Indeed, it has been shown that more accurate model estimates can
be obtained using inverse covariance estimation methods [21], [22] or graph learning methods, such as those introduced in our prior work [23], [24].

This paper proposes a novel graph-based modeling framework to design GBTs, where the models of interest are based on Gaussian-Markov random fields (GMRFs) whose inverse covariances are graph Laplacian matrices. The proposed framework consists of two distinct techniques to develop GBTs for video coding, called GL-GBTs and EA-GBTs:

- **Graph learning for GBT (GL-GBT) design:** A graph learning problem with a maximum-likelihood (ML) criterion is formulated and solved to estimate a graph Laplacian matrix from training data. In order to construct separable and nonseparable GBTs, two instances of the proposed problem with different connectivity constraints are solved by applying the graph learning algorithm introduced in our prior work [23]. Then, the GBTs are constructed by the eigendecomposition of the graph Laplacians of the learned graphs. As the KLT, a GL-GBT is learned from a sample covariance, but in addition, it incorporates Laplacian and structural constraints reflecting the inherent model assumptions about the video signal. From a statistical learning perspective, the main advantage of the proposed GL-GBT over the KLT is that the GL-GBT requires learning fewer model parameters from training data, and thus can lead to a more robust transform allowing better compression for the block signals outside of the training dataset. GL-GBTs can be adopted to improve existing MDT or RDOT schemes using KLTs.

- **Edge-adaptive GBT (EA-GBT) design:** To adapt transforms for block signals with image edges, we develop edge-adaptive GBTs (EA-GBTs) which are designed on a per-block basis. These lead to a block-adaptive transform scheme, where transforms are derived from graph Laplacians whose weights are modified based on image edges detected in a residual block.

In the literature, there are a few studies on model-based transform designs for image and video coding. In [23], [26], the authors present a graph-based probabilistic framework for predictive video coding and use that to justify the optimality of DCT, yet optimal graph/transform design is out of their scope. In our previous work [27], we present a comparison of various instances of different graph learning problems for nonseparable image modeling. The present paper theoretically and empirically validates one of the conclusions in [27], which suggests the use of GBTs derived from graph Laplacian matrices for image compression. In [28], a block-adaptive scheme is proposed for image compression with GBTs, where a graph is constructed for each block signal by minimizing a regularized Laplacian quadratic term used as the proxy for actual RD cost. On the other hand, in our present work, graphs are estimated from aggregate data statistics or constructed using image edge information. Then, the encoder selects the best transform based on exact RD measures. Moreover, Shen et. al. [29] propose edge adaptive transforms (EAT) specifically for depth-map compression, and Hu et. al. [30] extend EATs for piecewise-smooth image compression. Although our proposed EA-GBTs adopt some basic concepts originally introduced in [29], our graph construction method is different (in terms of image edge detection) and provides better compression for residual signals.

The main contributions of this paper can be summarized as follows:

- We propose graph learning techniques for separable and nonseparable GBTs, i.e., GBSTs and GBNTs, respectively, and present a theoretical justification of their use for coding residual block signals modeled using GMRFs.
- As an extension of the 1-D GMRF models used to design GBSTs for intra and inter predicted signals in our previous work [2], we present a general 2-D GMRF model for GBNTs and analyze its optimality.
- We apply EA-GBTs to intra and inter predicted blocks, while our prior work in [1] focused only on inter predicted blocks. In addition to the experimental results, we further derive some theoretical results and discuss the cases in which EA-GBTs are useful.
- We present comprehensive experimental results comparing the compression performances obtained using KLTs, GL-GBTs and EA-GBTs.

The rest of the paper is organized as follows. Section II presents the basic notation and definitions used throughout the paper. Section III introduces 2-D GMRFs used for modeling the video signals and discusses graph-based interpretations. In Section IV, the GBT design problem is formulated as a graph Laplacian estimation problem, and solutions for optimal GBNT and GBST construction are proposed. Section V presents EA-GBTs. Graph-based interpretations of residual block signal characteristics are discussed in Section VI by empirically validating the theoretical observations. Experimental results are presented in Section VII and Section VIII draws concluding remarks.

II. NOTATION AND PRELIMINARIES

Throughout the paper, lowercase normal (e.g., a and θ), lowercase bold (e.g., a and Θ) and uppercase bold (e.g., A and Θ) letters denote scalars, vectors and matrices, respectively. Unless otherwise stated, calligraphic capital letters (e.g., ℰ and ℳ) represent sets. Notation is summarized in Table 1.

In this paper, we focus on connected, undirected, weighted simple graphs with nonnegative edge weights [31]. We next present basic definitions related to graphs.

**Definition 1 (Weighted Graph).** The graph \( G = (V, ℰ, f_w, f_v) \) is a weighted graph with \( n \) vertices in the set \( V = \{ v_1, \ldots, v_n \} \). The edge set \( ℰ = \{ e \mid f_w(e) \neq 0, \forall e \in P_u \} \) is a subset of \( P_u \), the set of all unordered pairs of vertices, where \( f_w((v_i, v_j)) \geq 0 \) for \( i \neq j \) is a real-valued edge weight function, and \( f_v(v_i) \) for \( i = 1, \ldots, n \) is a real-valued vertex (self-loop) weight function.

**Definition 2 (Algebraic representations of graphs).** For a given weighted graph \( G = (V, ℰ, f_w) \) with \( n \) vertices, \( v_1, \ldots, v_n \):

The adjacency matrix of \( G \) is an \( n \times n \) symmetric matrix, \( W \), such that \( W_{ij} = W_{ji} = f_w((v_i, v_j)) \) for \( (v_i, v_j) \in P_u \).
The degree matrix of $\mathcal{G}$ is an $n \times n$ diagonal matrix, $D$, with entries $(D)_{ii} = \sum_{j=1}^{n} (W)_{ij}$ and $(D)_{ij} = 0$ for $i \neq j$.

The self-loop matrix of $\mathcal{G}$ is an $n \times n$ diagonal matrix, $V$, with entries $(V)_{ii} = f_\nu(v_i)$ for $i = 1, \ldots, n$ and $(V)_{ij} = 0$ for $i \neq j$. If $G$ is a simple weighted graph, then $V = 0$.

The connectivity matrix of $\mathcal{G}$ is an $n \times n$ matrix, $A$, such that $(A)_{ij} = 1$ if $(W)_{ij} \neq 0$, and $(A)_{ij} = 0$ if $(W)_{ij} = 0$ for $i, j = 1, \ldots, n$, where $W$ is the adjacency matrix of $\mathcal{G}$.

The combinatorial graph Laplacian (CGL) of graph $\mathcal{G}$ is defined as $L = D - W$.

The generalized graph Laplacian (GGL) of graph $\mathcal{G}$ is defined as $L = D - W + V$, which reduces to the combinatorial graph Laplacian when there are no self-loops ($V = 0$).

**Definition 3** (Graph-based Transform (GBT)). Let $L$ be a graph Laplacian of a graph $\mathcal{G}$. The graph-based transform is the orthogonal matrix $U$, satisfying $U^T U = I$, obtained by eigendecomposition of $L = U A U^T$, where $A$ is the diagonal matrix consisting of eigenvalues of $L$ (graph frequencies).

As formally stated in the following proposition, GTBs are invariant under (i) constant scaling of graph weights and (ii) addition of a constant self-loop weight to all vertices.

**Proposition 1.** Let $U$ be a GBT diagonalizing graph Laplacian $L$. The same $U$ also diagonalizes Laplacians of the form $L = c_1 L + c_2 I$, where $c_1$ and $c_2$ are real-valued scalars.

**Proof.** The proof is straightforward from Definition 3.

III. **Graph-based Models and Transforms for Video Block Signals**

For modeling video block signals, we use Gaussian Markov random fields (GMRFs), which provide a probabilistic interpretation for our graph-based framework. Assuming that the random vector of interest $x \in \mathbb{R}^n$ has zero mean\(^2\), a GMRF model for $x$ is defined based on a precision matrix $\Omega$, so that $x$ has a multivariate Gaussian distribution, $x \sim N(0, \Sigma^{-1})$,

$$p(x|\Omega) = \frac{1}{(2\pi)^{n/2}|\det(\Omega)|^{1/2}} \exp \left(-\frac{1}{2} x^T \Omega x \right).$$

with covariance matrix $\Sigma = \Omega^{-1}$. The entries of the precision matrix $\Omega$ can be interpreted in terms of conditional dependence relations among variables,

$$E \left[ x_i | (x)_{S\setminus{i}} \right] = -\frac{1}{\Omega_{ii}} \sum_{j \in S \setminus i} (\Omega)_{ij} x_j$$

$$\text{Prec} \left[ x_i | (x)_{S\setminus{i}} \right] = (\Omega)_{ii}$$

$$\text{Corr} \left[ x_i, x_j | (x)_{S\setminus{i, j}} \right] = -\frac{(\Omega)_{ij}}{\sqrt{(\Omega)_{ii} (\Omega)_{jj}}}$$

where $S = \{1, \ldots, n\}$ is the index set for $x = [x_1, \ldots, x_n]^T$.

The conditional expectation in (2) gives the best minimum mean square error (MMSE) estimate of $x_i$ using all other random variables. The relation in (3) corresponds to the precision of $x_i$ and (4) to the partial correlation between $x_i$ and $x_j$ (i.e., correlation between random variables $x_i$ and $x_j$ given all other variables in $x$). For example, if $x_i$ and $x_j$ are conditionally independent (i.e., $(\Omega)_{ij} = 0$), there is no edge between corresponding vertices $v_i$ and $v_j$. If all partial correlations are nonnegative (i.e., off-diagonal elements of $\Omega$ are nonpositive), then the model in (1) is classified as attractive GMRF\(^2\), whose precision matrix satisfies the following proposition 2.

**Proposition 2.** A GMRF model is attractive if and only if its precision matrix $\Omega$ is a generalized graph Laplacian matrix.

**Proof.** The proof is straightforward by the definitions.
In statistical modeling of image/video signals, it is generally assumed that adjacent pixel values are positively correlated [34], [35]. The assumption is intuitively reasonable for video signals, since neighboring pixel values are often similar to each other due to spatial and temporal redundancy. With this general assumption, we propose attractive GMRFs to model intra/inter predicted video block signals. Figs. 2 and 3 illustrate the 1-D and 2-D GMRFs defining line and grid graphs that are used to design separable and nonseparable GBTs, respectively. We formally define GBST and GBNT as follows.

**Definition 4** (Graph-based Separable Transform–GBST). Let \( U_{\text{row}} \) and \( U_{\text{col}} \) be \( N \times N \) GBTs associated with two line graphs with \( N \) vertices, then the GBST of an \( N \times N \) matrix \( X \) is

\[
\mathbf{X} = U_{\text{col}}^\intercal X U_{\text{row}},
\]

where \( U_{\text{row}} \) and \( U_{\text{col}} \) are rows and columns of the block signal \( X \), respectively.

**Definition 5** (Graph-based Nonseparable Transform–GBNT). Let \( U \) be an \( N^2 \times N^2 \) GBT associated with a graph with \( N^2 \) vertices, then the GBNT of an \( N \times N \) matrix \( X \) is

\[
\mathbf{X} = \text{block}(U^\intercal \text{vec}(X)),
\]

where \( U \) is applied on vectorized signal \( X = \text{vec}(X) \), and the block operator restructures the signal back in block form.

In our previous work [2], we introduced the 1-D GMRFs illustrated in Fig. 2 for intra/inter predicted signals and also derived closed-form expressions of their precision matrices (i.e., \( \Omega \)). The following section presents a unified 2-D extension of those models. While the work in [26] has noted the relation between graph Laplacians and GMRFs in the context of predictive transform coding, the following section further shows that residual signals obtained from MMSE prediction follow an attractive GMRF model. Hence, the optimal linear transform decorrelating such residual signals is a GBT derived from a GGL.

### A. 2-D GMRF Model for Residual Signals

We introduce a general 2-D GMRF model for intra/inter predicted \( N \times N \) block signals depicted in Fig. 3 by deriving the precision matrix of the residual signal \( r \), obtained after predicting the signal \( x = \left[ x_1 \ x_2 \ \cdots \ x_n \right]^\intercal \) with \( n = N^2 \) from \( n_p \) reference samples in \( y = \left[ y_1 \ y_2 \ \cdots \ y_{n_p} \right]^\intercal \) (i.e., predicting unfilled vertices from black filled vertices in Fig. 3), where \( x \) and \( y \) are zero-mean and jointly Gaussian with respect to the following attractive 2-D GMRF:

\[
p((x|y)\Omega = \frac{1}{(2\pi)^{n/2} \det(\Omega)^{1/2}} \exp \left(-\frac{1}{2} \left(\begin{array}{c} x \\ y \end{array} \right)^\intercal \Omega^{-1} \left(\begin{array}{c} x \\ y \end{array} \right) \right).
\]

The precision matrix \( \Omega \) and the covariance matrix \( \Sigma = \Omega^{-1} \) can be partitioned as follows [26], [33]:

\[
\Omega = \left[ \begin{array}{cc} \Omega_x & \Omega_{xy} \\ \Omega_{yx} & \Omega_y \end{array} \right] = \left[ \begin{array}{cc} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{array} \right]^{-1} = \Sigma^{-1}.
\]

Irrespective of the type of prediction (intra/inter), the MMSE prediction of \( x \) from the reference samples in \( y \) is

\[
p = E[x|y] = \Sigma_{xy} \Sigma_y^{-1} y = -\Omega_x \Sigma_{xy} y,
\]

and the resulting residual vector is \( r = x - p \) with covariance

\[
\Sigma_r = \Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx} - 2 \Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx} = \Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx}.
\]

By the matrix inversion lemma [36], the precision matrix of the residual \( r \) is shown to be equal to \( \Omega_x \), that is the submatrix in (8).

\[
\Sigma_r^{-1} = \left( \Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx} \right)^{-1} = \Omega_x.
\]

Since we also have \( \Sigma_x = (\Omega_x - \Omega_{xy} \Omega_y^{-1} \Omega_{yx}) \) by [36], the desired precision matrix can also be written as

\[
\Omega_{\text{residual}} = \Sigma_r^{-1} = \Sigma_x = \Sigma_x^{-1} + \Omega_{xy} \Omega_y^{-1} \Omega_{yx}.
\]

This construction leads us to the following proposition formally stating the conditions for a residual signal (i.e., \( r \)) to be modeled by an attractive GMRF.

**Proposition 3.** Let the signals \( x \in \mathbb{R}^n \) and \( y \in \mathbb{R}^{n_p} \) be distributed based on the attractive GMRF model in (7) with precision matrix \( \Omega \). If the residual signal \( r \) is estimated by minimum mean square error (MMSE) prediction of \( x \) from \( y \) (i.e., \( r = x - E[x|y] \)), then the residual signal \( r \) is distributed as an attractive GMRF whose precision matrix is a generalized graph Laplacian (i.e., \( \Omega_{\text{residual}} \) in (12)).

**Proof.** The proof follows from (7)–(12) where the inverse covariance of residual signal \( r \), \( \Sigma_r^{-1} \), is shown to be equal to \( \Omega_x \). Since \( \Omega_x \) is a submatrix of \( \Omega \) in (8) and \( \Omega \) is a GGL, \( \Omega_{\text{residual}} = \Omega_x \) is also a GGL. Hence, \( r \) is distributed as an attractive GMRF whose precision is \( \Omega_x \).

Note that Proposition 3 also applies to the 1-D signal models presented in [2] which are special cases of (7).

### B. Interpretation of Graph Weights for Predictive Transform Coding

Based on Proposition 3, the distribution of residual signals, denoted as \( r = \left[ r_1 \ \cdots \ r_n \right]^\intercal \), is defined by the following GMRF whose precision matrix is a GGL matrix \( L \) (i.e., \( L = \Omega_{\text{residual}} \)).

\[
p(r|L) = \frac{1}{(2\pi)^{n/2} \det(L)^{-1/2}} \exp \left(-\frac{1}{2} r^\intercal L r \right),
\]

where the quadratic term in the exponent can be decomposed in terms of graph weights (i.e., \( V \) and \( W \)) as

\[
r^\intercal L r = \sum_{i=1}^n (V_{ii}) r_i^2 + \sum_{(i,j) \in \mathcal{E}} (W_{ij}) (r_i - r_j)^2
\]

such that \( (W_{ij}) = -(L)_{ij} \), \( (V_{ii}) = \sum_{j=1}^n (L)_{ij} \), and \( \mathcal{E} = \{(i,j) \mid (v_i, v_j) \in \mathcal{E} \} \) is the set of index pairs of all vertices associated with the edge set \( \mathcal{E} \).

Based on (13) and (14), it is clear that the distribution of the residual signal \( r \) depends on edge weights (\( W \)) and vertex weights (\( V \)) where

- a model with larger (resp. smaller) edge weights (e.g., \( (W_{ij}) \)) increases the probability of having a smaller (resp. larger) squared difference between corresponding residual pixel values (e.g., \( r_i \) and \( r_j \)),
- a model with larger (resp. smaller) vertex weights (e.g., \( (V_{ii}) \)) increases the probability of pixel values (e.g., \( r_i \)) with smaller (resp. larger) magnitude.
In practice, a characterization of the edge and vertex weights ($W$ and $V$) can be made by estimating $L$ from data, which depend on inherent signal statistics and the type of prediction used for predictive coding. We empirically investigate the graph weights associated with residual signals in Section VII.

**C. DCTs/DSTs as GBTs Derived from Line Graphs**

Some types of DCTs and DSTs, including DCT-2 and DST-7, are in fact special cases of GBTs derived from Laplacians of specific line graphs. The relation between different types of DCTs and graph Laplacian matrices is originally discussed in [37] where DCT-2 is shown to be equal to the GBT uniquely obtained from graph Laplacians of the following form:

$$L_u = \begin{bmatrix} c & -c & 0 \\ -c & 2c & -c \\ \vdots & \vdots & \ddots \\ -c & 2c & -c \\ 0 & -c & c \end{bmatrix}$$

for $c > 0$, (15)



which represents uniformly weighted line graphs with no self-loops (i.e., all edge weights are equal to a positive constant and vertex weights to zero). Moreover, in [2], [38], it has been shown that DST-7 is the GBT derived from a graph Laplacian $L = L_u + V$ where $V = \text{diag}([c 0 \cdots 0]^T)$ including a self-loop at vertex $v_1$ with weight $f_v(v_1) = c$. Based on the results in [37], [39], various other types of DCTs and DSTs can be characterized using graphs. Table II specifies the line graphs (with $n$ vertices $v_1, v_2, \ldots, v_n$ having self-loops at $v_1$ and $v_n$) corresponding to different types of DCTs and DSTs, which are GBTs derived from graph Laplacians of the form $L = L_u + V$ where $V = \text{diag}([f_v(v_1) 0 \cdots 0 f_v(v_n)]^T)$.

| Vertex weights $f_v(v_1) = 0$ | $f_v(v_1) = c$ | $f_v(v_1) = 2c$ |
|-----------------------------|---------------|---------------|
| $f_v(v_n) = c$ | DCT-2 | DST-7 | DST-4 |
| $f_v(v_n) = 2c$ | DCT-8 | DST-1 | DST-6 |

**IV. GRAPH LEARNING FOR GRAPH-BASED TRANSFORM DESIGN**

**A. Generalized Graph Laplacian Estimation**

As justified in Proposition 3, the residual signal $r \in \mathbb{R}^n$ is modeled as an attractive GMRF, $r \sim N(0, \Sigma = L^{-1})$, whose precision matrix is a GGL denoted by $L$. Assuming that we have $k$ residual signals, $r_1, \ldots, r_k$, sampled from $N(0, \Sigma = L^{-1})$, the likelihood of a candidate $L$ is

$$\prod_{i=1}^k p(r_i|L) = (2\pi)^{-\frac{n}{2}} \det(L)^{-\frac{1}{2}} \prod_{i=1}^k \exp \left( -\frac{1}{2} r_i^T L r_i \right).$$

(16)

The maximization of the likelihood in (16) can be equivalently formulated as minimizing the negative log-likelihood, that is

$$\bar{L}_{\text{ML}} = \arg \min_L \left\{ \frac{1}{2} \sum_{i=1}^k \text{Tr} (r_i^T L r_i) - \frac{k}{2} \log \det(L) \right\}$$

(17)

where $S = \frac{1}{k} \sum_{i=1}^k r_i r_i^T$ is the sample covariance, and $\bar{L}_{\text{ML}}$ denotes the maximum likelihood (ML) estimate of $L$. To find the best GGL from a set of residual signals $\{r_1, \ldots, r_k\}$ in a maximum likelihood sense, we solve the following GGL estimation problem with connectivity constraints:

$$\begin{align*}
\text{minimize} & \quad \text{Tr} \left( L S \right) - \log \det(L) \\
\text{subject to} & \quad (L)_{ij} \leq 0 \text{ if } (A)_{ij} = 1 \\
& \quad (L)_{ij} = 0 \text{ if } (A)_{ij} = 0
\end{align*}$$

(18)

where $S$ denotes the sample covariance of residual signals, and $A$ is the connectivity matrix representing the graph structure (i.e., the set of graph edges). In order to optimally solve (18), we use the GGL estimation algorithm proposed in our previous work on graph learning [23].

**B. Graph-based Transform Design**

To design separable and nonseparable GBTs, we solve several instances of (18), each of which is denoted as GGL($S, A$). Then, the optimized GGL matrices are used to derive GBTs.

**Graph-based Separable Transforms (GBST).** For the GBST design, we solve two instances of (18) to optimize two separate line graphs used to derive $U_{\text{row}}$ and $U_{\text{col}}$ in [5]. Since we wish to design a separable transform, each line graph can be optimized independently. Thus, our basic goal is finding the best line graph pair based on sample covariance matrices $S_{\text{row}}$ and $S_{\text{col}}$ created from rows and columns of residual block signals. For $N \times N$ residual blocks, the proposed GBST construction has following steps:

1. Create the connectivity matrix $A_{\text{line}}$ representing a line graph structure with $n = N$ vertices as in Fig. 2.
2. Obtain two $N \times N$ sample covariances, $S_{\text{row}}$ and $S_{\text{col}}$, from rows and columns of size $N$, respectively, obtained from residual blocks in the dataset.
3. Solve instances of the problem in (18), GGL($S_{\text{row}}, A_{\text{line}}$) and GGL($S_{\text{col}}, A_{\text{line}}$), by using the GGL estimation algorithm [23] to learn Laplacians $L_{\text{row}}$ and $L_{\text{col}}$ representing line graphs, respectively.
4. Perform eigendecomposition on $L_{\text{row}}$ and $L_{\text{col}}$ to obtain GBTs, $U_{\text{row}}$ and $U_{\text{col}}$, which define the GBST as in [5].

**Graph-based Nonseparable Transforms (GBNT).** Similarly, for $N \times N$ residual block signals, we propose following steps to design a GBNT:

1. Create the connectivity matrix $A$ based on a desired graph structure. For example, $A$ can represent a grid graph with $n = N^2$ vertices as in Fig. 3.
2. Obtain $N^2 \times N^2$ sample covariance $S$ using residual block signals in the dataset (after vectorizing the block signals).
3. Solve the problem GGL($S, A$) by using the GGL estimation algorithm [23] to estimate a Laplacian $L$.
4. Perform eigendecomposition on $L$ to obtain the $N^2 \times N^2$ GBNT, $U$ defined in [6].

**C. Theoretical Justification for GL-GBTs**

It has been shown that KLT is optimal for transform coding of jointly Gaussian sources in terms of mean-square error

\footnote{Alternatively, joint optimization of the transforms associated with rows and columns has been recently proposed in [41], [42].}
(MSE) criterion under high-bitrate assumptions \cite{43,45}. Since GMRF models lead to jointly Gaussian distributions, the corresponding KLTs are optimal in theory. However, in practice, a KLT is obtained by eigendecomposition of the associated sample covariance, which has to be estimated from a training dataset where the number of data samples may not be sufficient to accurately recover the parameters. As a result, the sample covariance may lead a poor estimation of the actual model parameters \cite{20,21}. To improve estimation accuracy and alleviate overfitting, it is often useful to reduce the number of model parameters by introducing model constraints and regularization. From the statistical learning theory perspective \cite{46,47}, the advantage of our proposed GL-GBT over KLT is that KLT requires learning \(O(n^2)\) model parameters while GL-GBT only needs \(O(n)\), given the connectivity constraints in \cite{18}. Therefore, our graph learning approach provides better generalization in learning the signal model by taking into account variance-bias tradeoff. This advantage can also be justified based on the following error bounds characterized assumptions, the error bound for estimating \(\Sigma\) with \(\Sigma\) derived in \cite{48} is

\[
||\Sigma - S||_F = O\left(\sqrt{\frac{n^2 \log(n)}{k}}\right), \tag{19}
\]

while estimating the precision matrix \(\Omega\) by using the proposed graph learning approach leads to the following bound shown in \cite{21},

\[
||\Omega - L||_F = O\left(\sqrt{\frac{n \log(n)}{k}}\right), \tag{20}
\]

where \(L\) denotes the estimated GGL. Thus, in terms of the worst-case errors (based on Frobenius norm), the proposed method provides a better model estimation as compared to the estimation based on the sample covariance. Section VII empirically justifies the advantage of GL-GBT against KLT.

V. EDGE-ADAPTIVE GRAPH-BASED TRANSFORMS

The optimality of GL-GBTs relies on the assumption that the residual signal characteristics are the same across different blocks. However, in practice, video blocks often exhibit complex image edge structures that can degrade the coding performance when the transforms are designed from average statistics without any classification based on image edges. In order to achieve better compression for video signals with image edges, we propose edge-adaptive GBTs (EA-GBTs), designed on a per-block basis, by constructing a graph whose weights are determined based on the salient image edges in each residual block.

A. EA-GBT Construction

To design an EA-GBT for a residual block, we first detect image edges based on a threshold \(T_{\text{edge}}\) applied on gradient values, obtained using the Prewitt operator on the block. Then, edge weights of a predefined graph are modified according to the locations of detected image edges, and the resulting graph is used to derive the associated GBT. As depicted in Fig. 4 to construct a graph, we start with a uniformly weighted grid graph for which all edge weights are equal to a fixed operator on the block. Then, the detected image edges on a given residual block (Fig. 4b) are used to determine the collocated edges in the graph, and the corresponding weights are reduced as \(w_e = w_e/s_{\text{edge}}\) (Fig. 4c), where \(s_{\text{edge}} \geq 1\) is a parameter modeling the sharpness of image edges (i.e., the level of differences between pixel values in presence of an image edge). Thus, a larger \(s_{\text{edge}}\) leads to smaller weights on edges connecting pixels (vertices) with an image edge in between.

According to our simulations on residual data, coding gains are observed for \(s_{\text{edge}} > 10\), which empirically corresponds to residuals with an intensity difference of at least 12 (between pixels adjacent to an image edge). In our experiments, a conservative threshold of \(T_{\text{edge}} = 10\) is used for image edge detection, and the parameter \(s_{\text{edge}}\) is set to 10. Although the proposed EA-GBT design can be extended with multiple \(s_{\text{edge}}\) parameters, our experiments showed that such extensions do not provide a good rate-distortion tradeoff due to the additional signaling overhead. For efficient signaling of detected edges, we employ arithmetic edge coding (AEC) \cite{49}, a state-of-the-art binary edge-map codec.

From the compression perspective, the EA-GBT construction can also be viewed as a classification procedure, so that each residual block (e.g., in Fig. 4b) is assigned to a class of signals associated with an attractive GMRF, whose corresponding graph (i.e., GGL) is determined by \(s_{\text{edge}}\) and image edge detection based on \(T_{\text{edge}}\) (e.g., in Fig. 4c). By using attractive GMRFs, the following subsection theoretically validates our experimental observation of achieving coding gains for \(s_{\text{edge}} > 10\).

B. Theoretical Justification for EA-GBTs

We present a theoretical justification for advantage of EA-GBTs over KLTs. For the sake of simplicity, our analysis is based on 1-D models with a single image edge, whose the location \(l\) is uniformly distributed as

\[
P(l = j) = \begin{cases} \frac{1}{N-1} & \text{for } j = 1, \ldots, N-1 \\ 0 & \text{otherwise} \end{cases} \tag{21}
\]
where $N$ is the number of pixels (i.e., vertices) on the line graph depicted in Fig. 5. This construction leads to a Gaussian mixture distribution based on $M = N - 1$ attractive GMRFs,

$$p(x) = \sum_{j=1}^{M} p(l = j) N(0, \Sigma_j)$$  \hfill (22)

with $\Sigma_j$ denoting the covariance of the $j$-th attractive GMRF, whose corresponding graph has an image edge between pixels $v_j$ and $v_{j+1}$ as illustrated in Fig. 5. Since $x$ follows a Gaussian mixture distribution, the KLT obtained from the covariance of $x$ (which implicitly performs a second-order approximation of the distribution) is suboptimal in MSE sense \[50\]. Especially, with many possible image edge locations and different orientations, the underlying distribution may contain a large number of mixtures (i.e., a large $M$), which makes learning a model from average statistics inefficient. On the other hand, the proposed EA-GBT removes the uncertainty due to the random variable $l$ by detecting the location of the image edge in pixel (vertex) domain, and then constructing a GBT based on the detected image edge. Yet, EA-GBT requires allocating additional bits to represent the image edge (side) information, while KLT only allocates bits for coding transform coefficients.

To demonstrate the rate-distortion tradeoff between KLT and EA-GBT based coding schemes, we use classical rate-distortion theory results with high-bitrate assumptions \[43\]–\[45\], in which the distortion ($D$) can be written as a function of bitrate ($R$),

$$D(R) = \frac{N}{12} 2^{2R_d} 2^{-2R}$$  \hfill (23)

with

$$R = \frac{R}{N} \quad \text{and} \quad H_d = \frac{1}{N} \sum_{i=1}^{N} H_d(c_i)$$  \hfill (24)

where $R$ denotes the total bitrate allocated to code transform coefficients in $c = U^T x$, and $H_d(c_i)$ is the differential entropy of code transform coefficients ($c_i$). For EA-GBT, $R$ is allocated to code both transform coefficients ($R_{\text{coeff}}^{\text{EA-GBT}}$) and side information ($R_{\text{edge}}$), so we have

$$R = R_{\text{coeff}}^{\text{EA-GBT}} + R_{\text{edge}} = R_{\text{coeff}}^{\text{EA-GBT}} + \log_2(M)$$  \hfill (25)

while for KLT, the bitrate is allocated only to code transform coefficients ($R_{\text{coeff}}^{\text{KLT}}$), so that $R = R_{\text{coeff}}^{\text{KLT}}$. Fig. 6 shows the coding gain of EA-GBT over KLT for different sharpness parameters (i.e., $s_{\text{edge}}$) in terms of the following metric, called coding gain,

$$\text{cg}(D_{\text{EA-GBT}}, D_{\text{KLT}}) = 10 \log_{10} \left( \frac{D_{\text{EA-GBT}}}{D_{\text{KLT}}} \right)$$  \hfill (26)

where $D_{\text{EA-GBT}}$ and $D_{\text{KLT}}$ denote distortion levels measured at high-bitrate regime for EA-GBT and KLT, respectively. EA-GBT provides better compression for negative $\text{cg}$ values in Fig. 6, which appear when the sharpness of edges $s_{\text{edge}}$ is large (e.g., $s_{\text{edge}} > 10$).

Note that the distortion function in (23) is derived based on high-bitrate assumptions. To characterize rate-distortion tradeoff for different bitrates, we employ the reverse water-filling technique \[43\], \[51\] by varying the parameter $\theta$ to obtain rate and distortion measures as follows

$$R(D) = \sum_{i=1}^{N} \frac{1}{2} \log_2 \left( \frac{\lambda_i}{D_i} \right)$$  \hfill (27)

where $\lambda_i$ is the $i$-th eigenvalue of the signal covariance, and

$$D_i = \begin{cases} \lambda_i & \text{if } \lambda_i \geq \theta \\ \theta & \text{if } \theta < \lambda_i \end{cases}$$  \hfill (28)

so that $D = \sum_{i=1}^{N} D_i$.

Figure 7 illustrates the coding gain formulated in (26) achieved at different bitrates, where each curve correspond to a different $s_{\text{edge}}$ parameter. Similar to Fig. 6, EA-GBT leads to a better compression if the sharpness of edges $s_{\text{edge}}$ is large (e.g., $s_{\text{edge}} > 10$ for $R/N > 0.6$). At low-bitrates (e.g., $R/N < 0.6$), EA-GBT can perform worse than KLT for $s_{\text{edge}} = 20, 40$, yet EA-GBT outperforms as bitrate increases.

VI. RESIDUAL BLOCK SIGNAL CHARACTERISTICS AND GRAPH-BASED MODELS

In this section, we discuss statistical characteristics of intra and inter predicted residual blocks, and empirically justify

\[6\]
our theoretical analysis and observations in Section III. Our empirical results are based on residual signals obtained by encoding 5 different video sequences (City, Crew, Harbour, Soccer and Parkrun) using the HEVC reference software (HM-14) [6] at 4 differentQP parameters \( \{QP = \{22, 27, 32, 37\}\} \). Although the HEVC standard does not implement optimal MMSE prediction (which is the main assumption in Section III), it includes 35 intra and 8 inter prediction modes, which provide reasonably good prediction for different classes of block signals.

Figs. 8–13 depict statistical characteristics of 8 × 8 residual signals for a few intra and inter prediction modes. In these figures, sample variances of residuals and corresponding graph-based models are illustrated. Both grid and line graphs (with normalized edge and vertex weights) are estimated from residual data by solving the GGL estimation problem in (18) used for GBNT and GBST construction.

Naturally, residual blocks have different statistical characteristics depending on the type of prediction and the prediction mode. The sample variances shown in Figs. 8a–13a for different prediction modes lead us to the following observations:

- As expected, inter predicted blocks have smaller sample variance (energy) across pixels compared to intra predicted blocks, because inter prediction provides better prediction with larger number of reference pixels as shown in Fig. 4.
- In intra prediction, sample variances are generally larger at the bottom-right part of residual blocks, since reference pixels are located at the top and left of a block where the prediction is relatively better. This holds specifically for planar, DC and diagonal modes using pixels on both top and left as references for prediction.
- For some angular modes including intra horizontal/vertical mode, only left/top pixels are used as references. In such cases the sample variance gets larger as distance from reference pixels increases. Fig. 10a illustrates sample variances corresponding to the horizontal mode.
- In inter prediction, sample variances are larger around the boundaries and corners of the residual blocks mainly because of occlusions leading to partial mismatches between reference and predicted blocks.
- In inter prediction, PU partitions lead to larger residual energy around the partition boundaries as shown in Fig. 13a corresponding to horizontal PU partitioning \( (N \times 2N) \).

Moreover, inspection of the estimated graphs in Figs. 8–13 leads to following observations, which validate our theoretical analysis and justify the interpretation of model parameters in terms of graph weights discussed in Section III.

- Irrespective of the prediction mode/type, vertex (self-loop) weights tend to be larger for the pixels that are connected to reference pixels. Specifically, in intra prediction, graphs have larger vertex weights for vertices (pixels) located at the top and/or left boundaries of the block (Figs. 8–11), while the vertex weights are approximately uniform across vertices in inter prediction (Figs. 12 and 13).

VII. EXPERIMENTAL RESULTS

A. Experimental Setup

In our experiments, we generate two residual block datasets, one for training and the other for testing. The residual blocks are collected by using HEVC reference software (HM version 14) [6]. For the training dataset, residual blocks are obtained by encoding 5 video sequences, City, Crew, Harbour, Soccer and Parkrun, and for the test dataset, we use 5 different video sequences, BasketballDrill, BQMall, Mobcal, Shields and Cactus. The sequences are encoded using 4 different quantization parameters, \( QP = \{22, 27, 32, 37\} \), and transform block sizes are restricted to 4 × 4, 8 × 8 and 16 × 16. In both datasets, residual blocks are classified based on the side information provided by the HEVC encoder [6]. Specifically, intra predicted blocks are classified based on 35 intra prediction modes offered in HEVC. Similarly, inter predicted blocks are classified into 7 different classes using prediction unit (PU) partitions, such that 2 square PU partitions are grouped as one class and other 6 rectangular PU partitions determine other classes. Hence, we have 35 + 7 = 42 classes in total. For each class and block size, the optimal GBST, GBNT and KLT are designed using the residual blocks in training dataset, while EA-GBTs are constructed based on the detected image edges. The details of transform the construction are discussed in Sections IV and V.

5For 4 × 4 and 16 × 16 residual blocks, the structure of sample variances and graphs are quite similar to the ones in Figs. 8a–13a.
Fig. 8. For the planar mode in intra prediction (a) shows the estimated sample variances of $8 \times 8$ residual signals. In (b) and (c), edge and vertex weights are shown for grid and line graphs learned from residual data, respectively. Darker colors represent larger values.

Fig. 9. For the DC mode in intra prediction (a) shows the estimated sample variances of $8 \times 8$ residual signals. In (b) and (c), edge and vertex weights are shown for grid and line graphs learned from residual data, respectively. Darker colors represent larger values.

Fig. 10. For the horizontal mode in intra prediction (a) shows the estimated sample variances of $8 \times 8$ residual signals. In (b) and (c), edge and vertex weights are shown for grid and line graphs learned from residual data, respectively. Darker colors represent larger values.

Fig. 11. For the diagonal mode in intra prediction (a) shows the estimated sample variances of $8 \times 8$ residual signals. In (b) and (c), edge and vertex weights are shown for grid and line graphs learned from residual data, respectively. Darker colors represent larger values.
To evaluate the performance of transforms, we adopt the mode-dependent transform (MDT) and the rate-distortion optimized transform (RDOT) schemes. The MDT scheme assigns a single transform trained for each mode and each block size. In RDOT scheme, the best transform is selected from a set of transforms $\mathcal{T}$ by minimizing the rate-distortion cost $J(\lambda_{RD}) = D + \lambda_{RD} R$ [52] where the multiplier $\lambda_{RD} = 0.85 \times 2^{(Q_P - 12)/3}$ depends on $Q_P$ parameter. In our simulations, different transform sets are chosen for each mode (i.e., class) and block size. Specifically, the RDOT scheme selects either the DCT or the transform designed for each mode and block size pair, so that the encoder has two transform choices for each block. 

Note that, this requires the encoder to send one extra bit to identify the RD optimized transform at the decoder side. For EA-GBTs, the necessary graph (i.e., image edge) information is also sent by using the arithmetic edge encoder (AEC) [49]. After the transformation of a block, the resulting transform coefficients are uniformly quantized, and then entropy coded using arithmetic coding [53]. The compression performance is measured in terms of Bjontegaard-delta rate (BD-rate) [54].

### B. Compression Results

Table III presents the overall coding gains achieved by using KLTs, GBSTs and GBNTs with MDT and RDOT schemes for intra and inter predicted blocks. According to the results, GBNT outperforms KLT irrespective of the prediction type and coding scheme. Fig. 14 further demonstrates the advantage of proposed approach over KLT when fewer number of training samples are available, where the performance difference between GBNT and KLT is increased as the number of available training samples are reduced. Specifically, the BD-rate gap between GBNT and KLT increases from 0.7% to 1.5% when twenty-percent of the training data is used instead of the complete data. This validates our observation that the proposed graph learning method leads to a more robust transform and provides a better generalization than KLT. Table III also shows that GBNT performs substantially better than GBST for coding intra predicted blocks, while for inter blocks GBST performs slightly better than GBNT. This is because, inter predicted residuals tend to have a separable structure as shown in Figs. 12 and 13, yet intra residuals have more directional structures as shown in Figs. 8, 9 and 11 which are better captured by using non-separable transforms. Moreover, RDOT scheme significantly outperforms MDT.

Table IV compares the RDOT coding performance of KLTs, GBSTs and GBNTs on residual blocks with different prediction modes. In RDOT scheme the transform sets are $\mathcal{T}_{KLT} = \{DCT, KLT\}$, $\mathcal{T}_{GBST} = \{DCT, GBST\}$ and $\mathcal{T}_{GBNT} = \{DCT, GBNT\}$, which consist of DCT and a trained transform for each mode and block size. The results show that GBNT consistently outperforms KLT for all prediction modes. Similar to Table III, GBST provides slightly better
In this work, we discuss the class of transforms, called graph-based transforms (GBTs), with their applications to video compression. In particular, separable and nonseparable GBTs are introduced and two different design strategies are proposed. Firstly, the GBT design problem is posed as a graph learning problem, where we estimate graphs from data and the resulting graphs are used to define GBTs (GL-GBTs). Secondly, we propose edge-adaptive GBTs (EA-GBTs) which can be adapted on a per-block basis using side-information (image edges in a given block). We also give theoretical justifications for these two strategies and show that well-known transforms such as DCTs and DSTs are special cases of GBTs, and graphs can be used to design generalized (e.g., DCT-like or DST-like) separable transforms. Our experiments demonstrate that GL-GBTs can provide considerable coding gains with respect to standard transform coding schemes using DCT. In comparison with the Karhunen-Loeve transform (KLT), GL-GBTs are more robust and provide better generalization. Although coding gains obtained by including EA-GBTs in addition to GL-GBTs in the RDOT scheme are limited, using EA-GBTs only provides considerable coding gains over DCT.

VIII. CONCLUSION

compression compared to KLT and GBST. Also for angular modes (e.g., diagonal mode) in intra predicted coding, GBNT significantly outperforms GBST as expected.

Table [V] demonstrates the RDOT coding performance of EA-GBTs for different modes. As shown in the table, the contribution of EA-GBT within the transform set $T_{GL-GBT-EA-GBT} = \{DCT, GL-GBT, EA-GBT\}$ is limited to 0.3% for intra predicted coding, whereas it is approximately 0.8% for inter coding. On the other hand, if the transform set is selected as $T_{EA-GBT} = \{DCT, EA-GBT\}$ the contribution of EA-GBT provides considerable coding gains, which are approximately 0.5% for intra and 1% for inter predicted coding.

REFERENCES

[1] H. E. Egilmez, A. Said, Y.-H. Chao, and A. Ortega, “Graph-based transforms for inter predicted video coding,” in IEEE International Conference on Image Processing (ICIP), Sept 2015, pp. 3992–3996.
[2] H. E. Egilmez, Y. H. Chao, A. Ortega, B. Lee, and S. Yea, “GBST: Separable transforms based on line graphs for predictive video coding,” in 2016 IEEE International Conference on Image Processing (ICIP), Sept 2016, pp. 2375–2379.
[3] H. E. Egilmez, “Graph-based models and transforms for signal/data processing with applications to video coding,” Ph.D. dissertation, University of Southern California, 2017.
[4] Y.-H. Chao, “Compression of signals on graphs with application to image and video coding,” Ph.D. dissertation, University of Southern California, 2017.
[5] W. B. Pennebaker and J. L. Mitchell, JPEG Still Image Data Compression Standard, 1st ed. Norwell, MA, USA: Kluwer Academic Publishers, 1992.
[6] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
[7] D. Mukherjee, J. Bankoski, R. S. Bulté, A. Grange, J. Han, J. Koleszar, P. Wilkins, and Y. Xu, “The latest open-source video codec: VP9 – an overview and preliminary results,” in Proc. 30th Picture Coding Symp., San Jose, CA, Dec. 2013.
[8] J. Han, A. Saxena, V. Melkote, and K. Rose, “Jointly optimized spatial prediction and block transform for video and image coding,” IEEE Trans. Image Process., vol. 21, no. 4, pp. 1874–1884, Apr. 2012.
[9] Y. Ye and M. Karczewicz, “Improved h.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning,” in IEEE International Conference on Image Processing (ICIP), Oct 2008, pp. 2116–2119.
[10] S. Takamura and A. Shimizu, “On intra coding using mode dependent 2D-KLT,” in Proc. 30th Picture Coding Symp., San Jose, CA, Dec. 2013, pp. 137–140.
[11] A. Arrufat, P. Philippine, and O. Deforges, “Non-separable mode dependent transforms for intra coding in hevc,” in 2014 IEEE Visual Communications and Image Processing Conference, Dec 2014, pp. 61–64.
[12] F. Zou, O. Au, C. Pang, J. Dai, X. Zhang, and L. Fang, “Rate-distortion optimized transforms based on the lloyd-type algorithm for intra block coding,” IEEE Journal of Selected Topics in Signal Processing, vol. 7, no. 6, pp. 1072–1083, Dec 2013.
[13] X. Zhao, J. Chen, M. Karczewicz, L. Zhang, X. Li, and W. J. Chien, “Enhanced multiple transform for video coding,” in 2016 Data Compression Conference (DCC), March 2016, pp. 73–82.
[14] X. Zhao, J. Chen, A. Said, V. Seregin, H. E. Egilmez, and M. Karczewicz, “Nsst: Non-separable secondary transforms for next generation video coding,” in 2016 Picture Coding Symposium (PCS 2016), December 2016, pp. 1–1.
[15] A. Said, X. Zhao, M. Karczewicz, H. E. Egilmez, V. Seregin, and J. Chen, “Highly efficient non-separable transforms for next generation video coding,” in 2016 Picture Coding Symposium (PCS 2016), December 2016, pp. 1–1.
[16] X. Zhao, J. Chen, M. Karczewicz, A. Said, and V. Seregin, “Joint separable and non-separable transforms for next-generation video coding,” IEEE Transactions on Image Processing, vol. 27, no. 5, pp. 2514–2525, May 2018.
[17] T. Biaiek, V. Lorcy, and P. Philippine, “Transform competition for temporal prediction in video coding,” IEEE Transactions on Circuits and Systems for Video Technology, pp. 1–1, 2018.
[18] Y.-H. Chao, A. Ortega, and S. Yea, “Graph-based lifting transform for intra-predicted video coding,” in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2016, pp. 1–1.
[19] Y. H. Chao, H. E. Egilmez, A. Ortega, S. Yea, and B. Lee, “Edge adaptive graph-based transforms: Comparison of steptemp edge models for video compression,” in 2016 IEEE International Conference on Image Processing (ICIP), Sept 2016, pp. 1539–1543.
[20] I. M. Johnstone and A. Y. Lu, “On consistency and sparsity for principal components analysis in high dimensions,” Journal of the American Statistical Association, vol. 104, no. 486, pp. 682–693, 2009.
[21] P. Ravikumar, M. Wainwright, B. Yu, and G. Raskutti, “High dimensional covariance estimation by minimizing l1-penalized log-determinant divergence,” Electronic Journal of Statistics (EJS), vol. 5, pp. 935–980, 2011.
### TABLE IV
Comparison of KLT, GBST and GBNT for coding of different prediction modes in terms of BD-rate with respect to the DCT.
Smaller (negative) BD-rates mean better compression.

| Transform Set | Intra Prediction | Inter Prediction |
|---------------|------------------|-----------------|
|               | Planar | DC     | Diagonal | Horizontal | All modes | Square | Rectangular | All modes |
| $T_{\text{KLT}}$ | $-5.79$ | $-4.57$ | $-7.68$ | $-6.14$ | $-6.02$ | $-3.47$ | $-2.93$ | $-3.35$ |
| $T_{\text{GBST}}$ | $-5.45$ | $-4.12$ | $-3.32$ | $-6.45$ | $-4.61$ | $-3.95$ | $-3.25$ | $-3.89$ |
| $T_{\text{GBNT}}$ | $-6.27$ | $-5.04$ | $-8.74$ | $-6.53$ | $-6.70$ | $-3.84$ | $-3.18$ | $-3.68$ |

### TABLE V
The contribution of EA-GBTs in terms of BD-rate with respect to the DCT.

| Transform Set | Intra Prediction | Inter Prediction |
|---------------|------------------|-----------------|
|               | Planar | DC     | Diagonal | Horizontal | All modes | Square | Rectangular | All modes |
| $T_{\text{GL-GBT}}$ | $-6.27$ | $-5.04$ | $-8.74$ | $-6.53$ | $-6.70$ | $-3.95$ | $-3.25$ | $-3.89$ |
| $T_{\text{EA-GBT}}$ | $-0.51$ | $-0.47$ | $-0.69$ | $-0.66$ | $-0.54$ | $-1.01$ | $-0.73$ | $-0.93$ |
| $T_{\text{GL-GBT+EA-GBT}}$ | $-0.58$ | $-3.54$ | $-9.07$ | $-6.89$ | $-7.01$ | $-4.80$ | $-3.65$ | $-4.73$ |

[22] J. Friedman, T. Hastie, and R. Tibshirani, “Sparse inverse covariance estimation with the graphical lasso,” *Biostatistics*, vol. 9, no. 3, pp. 432–441, Jul. 2008.

[23] H. E. Egilmez, E. Pavez, and A. Ortega, “Graph learning from data under Laplacian and structural constraints,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 825–841, Sept 2017.

[24] H. E. Egilmez, E. Pavez, and A. Ortega, “Graph learning from filtered signals: Graph system and diffusion kernel identification,” *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 2, pp. 360–374, June 2019.

[25] C. Zhang and D. Florencio, “Analyzing the optimality of predictive transform coding using graph-based models,” *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 106–109, 2013.

[26] C. Zhang, D. Florencio, and P. Chou, “Graph signal processing - a probabilistic framework,” Microsoft Research, Tech. Rep. MSR-TR-2015-31, April 2015. [Online]. Available: http://research.microsoft.com/apps/pubs/default.aspx?id=243326.

[27] E. Pavez, H. E. Egilmez, Y. Wang, and A. Ortega, “GTT: Graph template transforms with applications to image coding,” in *Picture Coding Symposium (PCS)*, 2015, pp. 199–203.

[28] G. Fracastoro, D. Thanou, and P. Frossard, “Graph-based transform coding with application to image compression,” *CoRR*, vol. abs/1712.06393, 2017. [Online]. Available: https://arxiv.org/abs/1712.06393.

[29] G. Shen, W.-S. Kim, A. Narang, A. Ortega, J. Lee, and H. Wey, “Edge-adaptive transforms for efficient depth map coding,” in *Picture Coding Symposium (PCS)*, 2010, pp. 566–569.

[30] W. Hu, G. Cheung, A. Ortega, and O. C. Au, “Multiresolution graph fourier transform for compression of piecewise smooth images,” *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 419–433, Jan 2015.

[31] F. R. K. Chung, *Spectral Graph Theory*. USA: American Mathematical Society, 1997.

[32] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. Cambridge, MA, USA: The MIT Press, 2009.

[33] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*, ser. Monographs on Statistics and Applied Probability. London: Chapman & Hall, 2005, vol. 104.

[34] A. K. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1989.

[35] A. M. Tekalp, *Digital Video Processing*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall Press, 2015.

[36] M. A. Woodbury, *Inverting Modified Matrices*, ser. Statistical Research Group Memorandum Reports. Princeton, NJ: Princeton University, 1950, no. 42.

[37] G. Strang, “The discrete cosine transform,” *SIAM Rev.*, vol. 41, no. 1, pp. 335–147, Mar. 1999.

[38] W. Hu, G. Cheung, and A. Ortega, “Intra-prediction and generalized graph Fourier transform for image coding,” *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1913–1917, Nov 2015.

[39] M. Püschel and J. M. F. Moura, “Algebraic signal processing theory: 1-D space,” *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3586–3599, 2008.

[40] H. E. Egilmez, E. Pavez, and A. Ortega, “GLL: Graph Laplacian learning package, version 1.0,” [https://github.com/STAC-USC/Graph_Learning](https://github.com/STAC-USC/Graph_Learning) 2017.

[41] H. E. Egilmez, O. G. Guleryuz, J. Ehmann, and S. Yea, “Row-column transforms: Low-complexity approximation of optimal non-separable transforms,” in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 2385–2389.

[42] E. Pavez, A. Ortega, and D. Mukherjee, “Learning separable transforms by inverse covariance estimation,” in *2017 IEEE International Conference on Image Processing (ICIP)*, Sept 2017, pp. 1–1.

[43] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA, USA: Kluwer Academic Publishers, 1991.

[44] V. K. Goyal, “Theoretical foundations of transform coding,” *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 9–21, Sep 2001.

[45] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3rd ed. Academic Press, 2008.

[46] V. N. Vapnik, “An overview of statistical learning theory,” *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, Sep 1999.

[47] U. von Luxburg and B. Schölkopf, *Statistical Learning Theory: Models, Concepts, and Results*. Amsterdam, Netherlands: Elsevier North Holland, May 2011, vol. 10, pp. 651–706.

[48] R. Vershynin, “How close is the sample covariance matrix to the actual covariance matrix?” *Journal of Theoretical Probability*, vol. 25, no. 3, pp. 655–686, 2012.

[49] I. Daribo, D. Florencio, and G. Cheung, “Arbitrarily shaped motion prediction for depth video compression using arithmetic edge coding,” *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4696–4708, Nov 2014.

[50] M. Effros, H. Feng, and K. Zeger, “Suboptimality of the Karhunen-Loeve transform for transform coding,” *IEEE Transactions on Information Theory*, vol. 50, no. 8, pp. 1605–1619, Aug 2004.

[51] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 1991.

[52] A. Ortega and K. Ramchandran, “Rate-distortion methods for image and video compression,” *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 23–50, Nov. 1998.

[53] A. Said, “Introduction to Arithmetic Coding: theory and practice,” HP Labs, Tech. Rep. HPL-2004-76, 2004.

[54] G. Bjontegaard, “Calculation of average PSNR differences between RD-curves,” *ITU-T Q.6/SG16 VCEG-M33 Contribution*, vol. 48, Apr 2001.