Analyzing the Optimality of Predictive Transform Coding Using Graph-Based Models

Cha Zhang, *Senior Member, IEEE*, and Dinei Florêncio, *Senior Member, IEEE*

***Abstract*—In this letter, we provide a theoretical analysis of optimal predictive transform coding based on the Gaussian Markov random field (GMRF) model. It is shown that the eigen-analysis of the precision matrix of the GMRF model is optimal in decorrelating the signal. The resulting graph transform degenerates to the well-known 2-D discrete cosine transform (DCT) for a particular 2-D first order GMRF, although it is not a unique optimal solution. Furthermore, we present an optimal scheme to perform predictive transform coding based on conditional probabilities of a GMRF model. Such an analysis can be applied to both motion prediction and intra-frame predictive coding, and may lead to improvements in coding efficiency in the future.**

***Index Terms*—Gaussian Markov random field, graph-based models, predictive coding, transform coding.**

I. Introduction

PREDICTIVE transform coding (PTC), also known as hybrid predictive/transform coding, has been the cornerstone for modern image/video codecs. Given an image block to be encoded, PTC first finds the best prediction of the block from encoded contents. One then subtracts the prediction from the image block, and applies a transform such as the discrete cosine transform (DCT) on the residue to compact its energy for quantization and entropy coding. Early compression standards such as MPEG-1 and MPEG-2 adopted motion prediction as the key component to improve coding efficiency for image sequences. The more recent H.264 AVC [18] and the upcoming HEVC [16] further introduced intra-frame prediction, which significantly enhances the coding efficiency for intra frames and still images.

The Karhunen-Loève transform (KLT) is known to be the only transform that fully decorrelates the signal [6]. The DCT was originally introduced as an approximation for the KLT of a first-order stationary Markov sequence [1]. A more rigorous analysis [3] showed that the KLT converges to the DCT as the correlation coefficient ρ tends to 1. Compared with KLT, DCT is extremely efficient to compute, and provides satisfactory performance for image and video coding. However, the optimality of DCT was recently questioned by researchers under the context of intra-frame predictive coding [19], [5], [21]. In such a coding mode, the prediction is based on the pixels that have been encoded in the *same* image, by extrapolating reconstructed reference pixels in a fixed set of directions. There are a total of 9 intra prediction modes in H.264/AVC, and 35 intra prediction modes in HEVC. The encoder tests all of them and selects the best mode in rate/distortion tradeoff. Since the predictions are directional, different pixels in the predicted blocks have different correlation with their neighbors. Therefore, DCT is no longer a good approximation of the optimal transform for the residual signal.

In [19], Ye and Karczewicz proposed the use of mode-dependent separable transforms to approximate an ideal KLT for coding the intra prediction residual. A number of follow-up works have exploited the use of symmetry to reduce the number of mode-dependent transform matrices [4], [17]. Han *et al.* [5] used a separable first-order Gauss-Markov model for the image signal to show that for certain prediction modes, a hybrid cosine/sine transform will yield better performance for the residual. Saxena and Fernandes [13] and Yeo *et al.* [20] further extended Han's approach in that the hybrid transform can be applied to more prediction modes, and the prediction process is also modified for optimal results.

In this letter, we conduct a theoretical analysis on the optimal PTC based on the Gaussian Markov random field (GMRF) model for images. We show that the eigen-analysis of the GMRF model's precision matrix can lead to the optimal decorrelation transform for the image, which we term the *graph transform* (GT). For a particular 1-D first order GMRF, this eigen-analysis produces the well-known DCT. We further extend the result by showing that the 2-D DCT is also an optimum transform for a similar class of 2-D signals, although it is not unique in achieving optimal decorrelation. The *graph-based transforms* in [14], [10] is also a special case of the GT, for the case where pixels across an edge are considered uncorrelated, and connected pixels' correlation tends to one. Finally, the probabilistic view of GMRF allows us to find the optimal predictive transform based on conditional probabilities. Thanks to the generality of GMRF models, we are no longer limited to separable or first order approximations as was done in the literature [19], [5], [21]. Our analysis framework can be applicable to both motion prediction and intra-frame predictive coding. While this letter is mostly theoretical, we believe such an analysis significantly enhances our understanding of the DCT, the graph-based transform, and predictive coding.

II. The Gaussian Markov Random Field Model

We start the analysis by introducing the GMRF model for images [12]. A GMRF is a restrictive multivariate Gaussian distribution that satisfies additional conditional independence assumptions. We often use a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to represent the conditional independence assumption, where $\mathcal{V}$ represents the set of nodes in the graph, and $\mathcal{E}$ represents the set of edges. For instance, Fig. 1 shows a few graphs representing image models.

Manuscript received September 06, 2012; revised November 07, 2012; accepted November 19, 2012. Date of publication November 27, 2012; date of current version December 07, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Giuseppe Scarpa.

C. Zhang and D. Florêncio are with Microsoft Research, Redmond, WA 98052 USA (e-mail: chazhang@microsoft.com; dinei@microsoft.com).

Digital Object Identifier 10.1109/LSP.2012.2230165

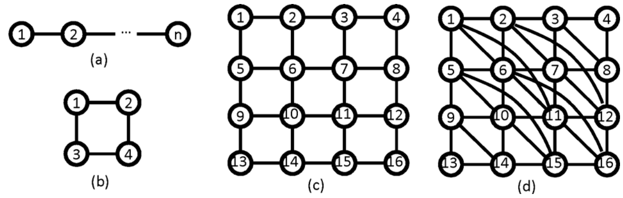

Fig. 1. A few graph models. (a) 1-D model, (b)–(c) simple 2-D models for a 2×2 and a 4×4 block, where each pixel is only connected with its direct neighbors, (d) a more complex model.

Pixels that are not directly connected will be conditionally independent given *all* other pixels.

Formally, a random vector $\mathbf{x} = (x_1,\ldots,x_n)^T$ is called a GMRF with respect to the graph $\mathcal{G} = (\mathcal{V} = \{1,\ldots,n\},\mathcal{E})$ with mean μ and a precision matrix $\mathbf{Q} > 0$ (positive definite), if and only if its density has the form [12]:

$$p(\mathbf{x}) = (2\pi)^{-\frac{n}{2}}|\mathbf{Q}|^{\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{Q}(\mathbf{x}-\mu)\right), \tag{1}$$

and

$$Q_{ij} \neq 0 \Leftrightarrow \{i,j\}\in\mathcal{E} \text{ for all } i\neq j. \tag{2}$$

In the above definition, the precision matrix $\mathbf{Q}$ is in fact the inverse of the covariance matrix $\mathbf{\Sigma}$ in a typical multivariate Gaussian distribution. It is sometimes beneficial to describe the Gaussian distribution using the precision matrix, as it can handle distributions with singular covariances. More importantly, partial correlations between variables can be directly obtained from $\mathbf{Q}$, since [12]:

$$\rho(x_i,x_j|\mathbf{x}\setminus\{x_i,x_j\}) = -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}},\quad i\neq j, \tag{3}$$

where $\rho(x_i,x_j|\mathbf{x}\setminus\{x_i,x_j\})$ represents the partial correlation between x_i and x_j given all other variables.

Although the above Gaussian model may be an idealized one when describing real-world images, similar models have been adopted to analyze image properties with great success [8]. In addition, GMRF has been widely used as an image prior model in various applications, such as image reconstruction [9], texture modeling and discrimination [2], segmentation [7], etc.

III. Graph Transform

A. The Graph Transform

Let us assume that an image follows a zero-mean, GMRF model. According to (1), the image vector $\mathbf{x}$ has a covariance matrix $\mathbf{\Sigma} = \mathbf{Q}^{-1}$, where $\mathbf{Q}$ is the precision matrix. The optimal linear transform that decorrelates $\mathbf{x}$ is thus the Karhunen-Loève transform, or the eigenvector matrix $\mathbf{\Phi}$ of $\mathbf{\Sigma}$. That is, we may write:

$$\mathbf{\Sigma\Phi} = \mathbf{\Phi\Lambda}, \tag{4}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_0,\ldots,\lambda_n)$ is the diagonal matrix of eigenvalues for $\mathbf{\Sigma}$. Since:

$$\begin{aligned}\mathbf{Q\Phi} &= \mathbf{\Sigma}^{-1}\mathbf{\Phi} = (\mathbf{\Phi\Lambda\Phi}^T)^{-1}\mathbf{\Phi}\\ &= \mathbf{\Phi\Lambda}^{-1},\end{aligned} \tag{5}$$

$\mathbf{\Phi}$ is also the eigenvector matrix of $\mathbf{Q}$.

Therefore, given a zero-mean GMRF with precision matrix $\mathbf{Q}$, the optimal decorrelation linear transform is the eigenvector matrix of the precision matrix. Since GMRF is defined on a graph, we call such an optimal transform the *Graph Transform*.

Note that GMRF is a very generic image model, since the precision matrix $\mathbf{Q}$ can be defined based on actual image statistics. Nevertheless, in the literature, the Laplacian matrix $\mathbf{L}$ of the graph $\mathcal{G}$ is most widely used as the image prior model for various applications. To compute the Laplacian $\mathbf{L}$, one first defines an adjacency matrix $\mathbf{A}$, where $\mathbf{A}(i,j) = \mathbf{A}(j,i) = 1$ if nodes i and j are immediate neighbors connected by an edge. Otherwise, $\mathbf{A}(i,j) = \mathbf{A}(j,i) = 0$. Alternatively, if there are weights on the graph that represent the partial correlation between nodes, $\mathbf{A}$ can be defined accordingly using the specified weights.

Further, the degree matrix $\mathbf{D}$ of graph $\mathcal{G}$ is defined as a diagonal matrix, whose ith diagonal element is the number of non-zero entries in the ith row of $\mathbf{A}$. The Laplacian of the graph is computed as:

$$\mathbf{L} = \mathbf{D} - \mathbf{A}. \tag{6}$$

The precision matrix $\mathbf{Q}$ of a GMRF can be defined as:

$$\mathbf{Q} = \delta\mathbf{L}, \tag{7}$$

where δ is a scale factor. Consequently, if the image follows such a GMRF model, the optimal linear decorrelation transform is the eigenvector matrix of the Laplacian matrix $\mathbf{L}$.

A transform design referred to as an edge adaptive transform (EAT) was recently proposed in depth map coding [14], where a graph is defined on image blocks, and the correlation across edges in the graph is set to 0. Our analysis above shows that under the GMRF model, the EAT is indeed the optimal transform to decorrelate the signal.

B. Graph Transform and 2-D DCT

The optimality of the 1-D DCT has been shown in the literature for an autoregressive signal model [3]. For the 1-D signal graph as shown in Fig. 1(a), it was also known that the eigenvector matrix of the Laplacian matrix $\mathbf{L}$ is identical to the discrete cosine transform (more specifically, DCT-2) [15]. While autoregressive model could be extended to 2-D images [11], to the best of our knowledge, the optimality of 2-D DCT has not been analyzed. In the following, we show that for 2-D images with graphs defined as Fig. 1(b) and (c), while the eigenvector matrix of the Laplacian may not be unique, the 2-D DCT basis functions are indeed eigenvectors of the graph Laplacian, and thus optimal.

Consider an image block of $M\times N$ pixels. The 2-D DCT basis functions of the image block can be written as a large $MN\times MN$ matrix $\mathbf{C}$, where each element is defined as:

$$c(i,j) = \alpha(k)\beta(l)\cos\left[\frac{\pi(2m+1)k}{2M}\right]\cos\left[\frac{\pi(2n+1)l}{2N}\right], \tag{8}$$

where $i = kN + l, j = mN + n$ for $0\leq k,m\leq M-1$ and $0\leq l,n\leq N-1$; $\alpha(0) = \sqrt{1/M}, \alpha(k) = \sqrt{2/M}$, for $1\leq k\leq M-1$; and $\beta(0) = \sqrt{1/N}, \beta(l) = \sqrt{2/N}$, for $1\leq l\leq N-1$.

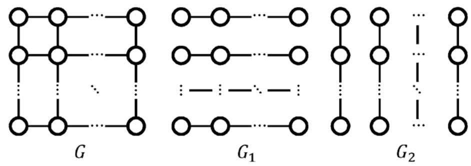

Fig. 2. Graph models on 2-D images. In graph $\mathcal{G}$ all pixels are connected to their four direct neighbors. In graph $\mathcal{G}_1$ the pixels are only connected to their horizontal neighbors. In graph $\mathcal{G}_2$ the pixels are only connected to their vertical neighbors.

Let us now define a graph on the image, with each pixel connecting to only its four direct neighbors, as the graph $\mathcal{G}$ in Fig. 2. Denote the Laplacian of the graph as $\mathbf{L}$. It can be verified, using examples such as the 2×2 image graph in Fig. 1(b) and 4×4 image graph in Fig. 1(c), that $\mathbf{L}$ may contain duplicated eigenvalues. Consequently, the optimal graph transform is not unique.

To show that the 2-D DCT matrix $\mathbf{C}$ is an eigenvector matrix for $\mathbf{L}$, we first define two assistant graphs, $\mathcal{G}_1$ and $\mathcal{G}_2$, as shown in Fig. 2. In $\mathcal{G}_1$ the pixels are only connected to their horizontal neighbors, while in $\mathcal{G}_2$ the pixels are only connected to their vertical neighbors. Their Laplacian matrices are $\mathbf{L}_1$ and $\mathbf{L}_2$, respectively. Due to their repeat structure, clearly neither $\mathbf{L}_1$ nor $\mathbf{L}_2$ have unique eigenvector matrices. The 2-D DCT is, however, an eigenvector matrix of both $\mathbf{L}_1$ and $\mathbf{L}_2$. The proof is a tedious but straightforward multiplication of $\mathbf{C}$ (as defined in (8)) by each of the Laplacian matrices, $\mathbf{L}_1$ and $\mathbf{L}_2$. Now, since:

$$\mathbf{L}=\mathbf{L}_1+\mathbf{L}_2, \tag{9}$$

and since $\mathbf{C}$ is an eigenvector matrix for both $\mathbf{L}_1$ and $\mathbf{L}_2$, it follows that it is also an eigenvector matrix for $\mathbf{L}$.

Thus, although the 2-D DCT is generally viewed simply as a computationally efficient extension of the 1-D DCT into 2-D, it turns out the 2-D DCT is actually optimum for a very reasonable signal model: the first order GMRF when the correlation coefficients goes to one. This confirms the successful application of the 2-D DCT in typical image coding designs such as JPEG. However, prediction is a key in more recent codecs, and that model is not as appropriate anymore. In the next section, we further extend our analysis, and examine *predictive* graph transform, where some pixels are predicted from known pixels, followed by an optimal graph transform.

IV. Predictive Graph Transform

A. Predictive Graph Transform

Let us consider a random vector $\mathbf{x}=(x_1,\ldots,x_n,x_{n+1},\ldots,x_m)^T$, and assume it follows a GMRF model with mean μ and precision matrix $\mathbf{Q}$. Among the elements of $\mathbf{x}$, $\mathbf{x}_1=(x_1,\ldots,x_n)^T$ is unknown, and $\mathbf{x}_2=(x_{n+1},\ldots,x_m)^T$ is known. For instance, $\mathbf{x}_1$ may represent the pixels to be encoded (unknown to the decoder), and $\mathbf{x}_2$ may represent a matching block in a previous frame during motion prediction, or a few nearby known pixels during intra-frame prediction.

The precision matrix $\mathbf{Q}$ can be partitioned as:

$$\mathbf{Q}=\begin{pmatrix}\mathbf{Q}_{11} & \mathbf{Q}_{12}\\ \mathbf{Q}_{12}^T & \mathbf{Q}_{22}\end{pmatrix}, \tag{10}$$

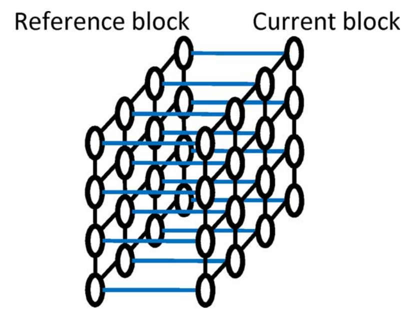

Fig. 3. An illustrative example for motion prediction. A 3D graph can be defined on the pixels to represent an example GMRF model.

where $\mathbf{Q}_{11}$ is mostly related to $\mathbf{x}_1$, $\mathbf{Q}_{22}$ is mostly related to $\mathbf{x}_2$, and $\mathbf{Q}_{12}$ represents the relationship between $\mathbf{x}_1$ and $\mathbf{x}_2$. According to [12], $\mathbf{x}_1|\mathbf{x}_2$ is also a GMRF with respect to its own subgraph with mean $\mu_{\mathbf{x}_1|\mathbf{x}_2}$ and precision matrix $\mathbf{Q}_{\mathbf{x}_1|\mathbf{x}_2}>0$, where:

$$\mu_{\mathbf{x}_1|\mathbf{x}_2}=\mu_{\mathbf{x}_1}-\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}(\mathbf{x}_2-\mu_{\mathbf{x}_2}), \tag{11}$$

$$\mathbf{Q}_{\mathbf{x}_1|\mathbf{x}_2}=\mathbf{Q}_{11}, \tag{12}$$

where $\mu_{\mathbf{x}_1}$ and $\mu_{\mathbf{x}_2}$ are the mean for $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. Consequently, in order to optimally decorrelate the variable $\mathbf{x}_1$ given $\mathbf{x}_2$, we can first subtract the conditional mean $\mu_{\mathbf{x}_1|\mathbf{x}_2}$ from $\mathbf{x}_1$, and then apply the eigenvector matrix of $\mathbf{Q}_{11}$ to transform the signal for further processing/compression.

In the following, we discuss the application of predictive graph transform (PGT) in motion prediction and intra-frame predictive coding. In both cases, we assume a generic GMRF model of the image is given during the analysis.

B. PGT for Motion Prediction

In motion prediction, a reference block is found through various motion estimation approaches, and is used to predict the block that is currently being encoded. Assume the two blocks are zero mean, and follow GMRF models described by precision matrix $\mathbf{Q}_{\mathrm{ref}}$ and $\mathbf{Q}_{\mathrm{c}}$, respectively. To this end, let us construct a GMRF model in 3D, as shown in Fig. 3. If we assume all "prediction" edges have weight one (since the reference block should be very similar to the current block due to motion search), the precision matrix of the 3D GMRF model can be written as:

$$\mathbf{Q}=\begin{pmatrix}\mathbf{Q}_{11} & -\mathbf{I}\\ -\mathbf{I} & \mathbf{Q}_{22}\end{pmatrix}, \tag{13}$$

where $\mathbf{I}$ is an identity matrix, and:

$$\mathbf{Q}_{11}=\mathbf{Q}_{\mathrm{c}}+\mathbf{I},\quad \mathbf{Q}_{22}=\mathbf{Q}_{\mathrm{ref}}+\mathbf{I}. \tag{14}$$

Based on the derivation in Section IV-A, we may predict the current block $\mathbf{x}_1$ through:

$$\mu_{\mathbf{x}_1|\mathbf{x}_2}=\mathbf{Q}_{11}^{-1}\mathbf{x}_2, \tag{15}$$

and then apply the eigenvector matrix of $\mathbf{Q}_{11}$ to decorrelate the signal.

The above analysis has interesting implications. For motion prediction, instead of directly copying the pixels from the reference block to the current block, (15) suggests that the optimal

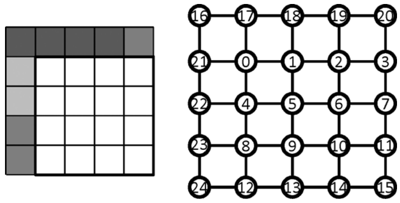

Fig. 4. An illustrative example for intra-frame prediction. The 4×4 image block will be predicted by the shaded known pixels on the top and left. The right figure is a typical graph defined on the image.

scheme is to first apply a filter on $\mathbf{x}_2$ before copying. In addition, since any orthogonal basis is an eigenvector matrix of the identity matrix $\mathbf{I}$, it can be shown that $\mathbf{Q}_{11}$ will share the same set of eigenvectors as $\mathbf{Q}_c$. Hence the optimal transform for the residue remains the same as when no motion prediction is performed.

For the special case that $\mathbf{Q}_c = \delta\mathbf{L}$, since the Laplacian is essentially a high pass filter, $\mathbf{Q}_{11}^{-1}$ will be a low-pass filter. Therefore, one should blur the reference block and then copy it to the current block. Furthermore, from the analysis in Section III-B, we can conclude that the 2-D DCT transform is still optimal for encoding the prediction residual.

C. PGT for Intra Predictive Coding

In modern video codecs, the intra frames will also be predicted from neighboring known pixels to enhance coding efficiency. Again we may form a simple graph for the 2-D block including the neighboring pixels, as shown in Fig. 4. Following Section IV-A, for a zero mean image, the optimal prediction would be:

$$\mu_{\mathbf{x}_1|\mathbf{x}_2} = -\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\mathbf{x}_2, \tag{16}$$

where $\mathbf{x}_1$ is the list of pixels to be encoded, and $\mathbf{x}_2$ is the list of known neighbors. The optimal transform is the eigenvector matrix of $\mathbf{Q}_{11}$.

Note that the optimal prediction for intra-frame predictive coding is related to both $\mathbf{Q}_{11}$ and $\mathbf{Q}_{12}$. That is, how the unknown pixels are correlated to themselves, and how they are correlated to the known pixels. In general the 2-D DCT is no longer the eigenvector matrix for $\mathbf{Q}_{11}$ due to the connections between $\mathbf{x}_1$ and $\mathbf{x}_2$. Similar to the previous works [19], [5], [13], our analysis calls for different schemes of intra-prediction and transform coding. On the other hand, our derivation is rather general, and not limited to separable or first order signal models.

If we consider the special case that $\mathbf{Q} = \delta\mathbf{L}$, both $\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}$ and the eigenvector matrix of $\mathbf{Q}_{11}$ can be pre-computed. In practice, however, the neighboring known pixels may suggest a better GMRF model, and it could certainly be adopted to improve the coding efficiency.

V. Conclusion

We presented a theoretical analysis of optimal predictive transform coding. We showed that, under a GMRF image model, the graph transform is optimal in decorrelating the signal. Furthermore, when neighboring pixels are known, we derived the optimal predictive graph transform as in (11) and (12). Mostly of historical interest, we have also showed that the 2-D DCT is optimum for a very reasonable image model, which helps understand its efficiency (and success) in early coding standards.

Note that throughout our predictive transform coding analysis, we assume that the GMRF image model is given and fixed. Predictive graph transform is optimal as long as the image follows the GMRF model closely. It is easy to imagine, however, that in many cases a more elaborate model can be devised after part (or all) of the data is available. This is precisely what is exploited in the directional interpolation in recent codecs. An interesting line of future work is to find better GMRF models based on the reference blocks or pixels. Together with the optimum predictive graph transform proposed in this letter, it would bring motion compensation and directional interpolation to a whole new level.

References

[1] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, no. 1, pp. 90–93, Jan. 1974.
[2] R. Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 33, no. 4, pp. 959–963, 1985.
[3] R. J. Clarke, *Transform Coding of Images*. New York: Academic, 1985.
[4] H. Yang, J. Zhou, and H. Yu, "Simplified MDDT (SMDDT) for intra prediction residual," in *Doc. JCTVC-B039, MPEG-H/JCT-VC*, 2010.
[5] J. Han, A. Saxena, and K. Rose, "Towards jointly optimal spatial prediction and adaptive transform in video/image coding," in *ICASSP*, 2010.
[6] A. K. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ: Prentice-Hall, 1989.
[7] S. Krishnamachari and R. Chellappa, "Multiresolution Gauss-Markov random field models for texture segmentation," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 251–267, 1997.
[8] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.
[9] E. Levitan and G. T. Herman, "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. Med. Imag.*, vol. 6, no. 3, pp. 185–192, 1987.
[10] E. Martinez-Enriquez, E. Diaz-de Maria, and A. Ortega, "Video encoder based on lifting transforms on graphs," in *ICIP*, 2011.
[11] J.-R. Ohm, *Multimedia Communication Technology*. New York: Springer, 2004.
[12] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*. London, U.K.: Chapman & Hall/CRC, 2005.
[13] A. Saxena and F. C. A. Fernandes, "Jointly optimal intra prediction and adaptive primary transform," in *Doc. JCTVC-C108, MPEG-H/JCT-VC*, 2010.
[14] G. Shen, W.-S. Kim, S. K. Narang, A. Ortega, J. Lee, and H. Wey, "Edge-adaptive transforms for efficient depth map coding," in *PCS*, 2010.
[15] G. Strang, "The discrete cosine transform," *SIAM Rev.*, vol. 41, no. 1, pp. 135–147, 1999.
[16] G. 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.*, Dec. 2012.
[17] A. Tanizawa, J. Yamaguchi, T. Shiodera, T. Chujoh, and T. Yamakage, "Improvement of intra coding by bidirectional intra prediction and 1 dimensional directional unified transform," in *Doc. JCTVC-B042, MPEG-H/JCT-VC*, 2010.
[18] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, 2003.
[19] Y. Ye and M. Karczewicz, "Improved H.264 intra coding based on bidirectional intra prediction, directional transform, and adaptive coefficient scanning," in *ICIP*, 2008.
[20] C. Yeo, Y. H. Tan, Z. Li, and S. Rahardja, "Mode-dependent fast separable KLT for block-based intra coding," in *ISCAS*, 2011.
[21] C. Yeo, Y. H. Tan, Z. Li, and S. Rahardja, "Mode-dependent transforms for coding directional intra prediction residuals," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 545–554, 2012.