

GRAPH-BASED ROTATION OF THE DCT BASIS FOR MOTION-ADAPTIVE TRANSFORMS

Du Liu and Markus Flierl

School of Electrical Engineering
KTH Royal Institute of Technology, Stockholm
{dul, mflierl}@kth.se

ABSTRACT

In this paper, we consider motion-adaptive transforms that are based on vertex-weighted graphs. The graphs are constructed by motion vector information and the weights of the vertices are given by scale factors, where the scale factors are used to control the energy compaction of the transform. The vertex-weighted graph defines a one dimensional linear subspace. Thus, our transform basis is subspace constrained. To find a full transform matrix that satisfies our subspace constraint, we rotate the discrete cosine transform (DCT) basis such that the first basis vector matches the subspace constraint. Since rotation is not unique in high dimensions, we choose a simple rotation that only rotates the DCT basis in the plane which is spanned by the first basis vector of the DCT and the subspace constraint. Experimental results on energy compaction show that the motion-adaptive transform based on this rotation is better than the motion-compensated orthogonal transform based on hierarchical decomposition while sharing the same first basis vector.

Index Terms— Motion-adaptive transform, subspace-constrained transform, vertex-weighted graph

1. INTRODUCTION

Our goal is to improve motion-compensated predictive coding by developing motion-adaptive transforms for efficient video coding. Since pixels in video sequences have high temporal and spatial correlation, we can form graphs for efficient linear transforms. [1] considers lifting transforms on graphs based on spatial and temporal connections due to block motion. [2] considers an edge-adaptive graph-based transform. The transform is constructed by the eigenvectors of the Laplacian matrix of the graph [3]. This method requires additional information on the edge map. In [4], wavelet transforms on graphs are proposed. The wavelets are based on the spectral decomposition of the discrete Laplacian matrix of the graph. One drawback is that the wavelet is overcomplete since there is no downsampling as in the classic discrete wavelet transform.

The so-called motion-compensated lifting wavelet permits a reversible filter structure, but struggles with unconnected, connected, and multi-connected pixels when performing the update step. Hence, [5] proposes an optimum update step that minimizes the mean-squared reconstruction error. On the other hand, the class of motion-compensated orthogonal transforms (MCOT) is designed for successive pictures in a video sequence and maintains strict orthogonality for any motion field [6, 7]. The transform compacts the signal energy efficiently by using the concept of scale factors. Further, sub-pel accurate motion and variable block sizes can be used with MCOT for efficient video coding [8].

In [9], we designed a class of motion-adaptive transforms based on vertex-weighted graphs. The vertex-weighted graph is constructed from the motion information only. Since the transform basis is derived from the eigen-decomposition of a subspace representation of the covariance matrix, the transform achieves optimal energy compaction given the subspace constraint. The advantage is that the constraint is defined by the vertex-weighted graph only. One disadvantage of the method is that many basis vectors are signal dependent. Thus, they need to be signaled to the decoder.

In this paper, we design a transform that takes the same vertex-weighted graph into consideration, but the basis vectors are not signal dependent. Our transform can be obtained simply by rotating the discrete cosine transform basis. Given the same vertex-weighted graph, it is designed to achieve better energy compaction than MCOT.

The paper is organized as follows: Section 2 defines the vertex-weighted graph. Section 3 discusses the construction of the transform basis by using a graph-based rotation. Section 4 presents the experimental results.

2. SUBSPACE CONSTRAINT

We consider in [9] a subspace-constrained transform that is based on vertex-weighted graphs. The graph is constructed by motion vector information. The vertex weights are given by the scale factors, which are defined according to [6]. As the vertex-weighted graph determines the transform in a one-dimensional linear subspace, the transform is subspace constrained. In the following, we explain the scale factors and

This work has been supported in part by the Swedish Research Council under the grant 2011-5841.

the subspace constraint. Our graph-based rotation also satisfies this subspace constraint.

Let $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ be a vector representing n connected pixels. Energy compaction changes the magnitude of the pixel intensities, while the model of ideal motion implies constant intensity for connected pixels. In order to combine energy compaction and ideal motion, we use the concept of scale factors from [6] and represent the original pixel values by x'_1, x'_2, \dots, x'_n . Then, let c_1, c_2, \dots, c_n be the scale factors of the pixels x_1, x_2, \dots, x_n , respectively, i.e., $x_k = c_k x'_k$, $k = 1, 2, \dots, n$. Similar to [6], we assume that these n pixels are connected by ideal motion such that $x'_1 = x'_2 = \dots = x'_n$. The weights of the vertices are then the scale factors and the vector of scale factors is $\mathbf{c} = [c_1, c_2, \dots, c_n]^T$. Then, an orthonormal transform matrix $T = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_n]$ compacts all the energy into the first coefficient,

$$\mathbf{y} = T^T \mathbf{x} = T^T \begin{bmatrix} c_1 x'_1 \\ c_2 x'_1 \\ \vdots \\ c_n x'_1 \end{bmatrix} = \begin{bmatrix} \sqrt{\sum_{k=1}^n c_k^2} \cdot x'_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (1)$$

where \mathbf{y} is the output of the transform. (1) can be rewritten as

$$\mathbf{t}_1^T \mathbf{c} = \sqrt{\mathbf{c}^T \mathbf{c}}, \quad (2)$$

$$\mathbf{t}_k^T \mathbf{c} = 0, \quad k = 2, 3, \dots, n. \quad (3)$$

That is, \mathbf{t}_1 captures the whole energy of the signal and $\mathbf{t}_2, \dots, \mathbf{t}_n$ capture no energy. Considering an n -dimensional orthogonal space defined by n orthonormal vectors $\mathbf{t}_1, \dots, \mathbf{t}_n$, the above two equations (2) and (3) are projections of \mathbf{c} onto this n -dimensional orthogonal space. The first vector \mathbf{t}_1 should be collinear with \mathbf{c} such that the projection of \mathbf{c} onto \mathbf{t}_1 contains the total energy. Since the space is orthogonal, no energy is left to be projected on the remaining $(n-1)$ -dimensional subspace. Thus, \mathbf{t}_1 is determined by \mathbf{c} , which gives our subspace constraint

$$\mathbf{t}_1 = \frac{\mathbf{c}}{\|\mathbf{c}\|_2} = \frac{\mathbf{c}}{\sqrt{\mathbf{c}^T \mathbf{c}}}. \quad (4)$$

If \mathbf{x} is not due to ideal motion, i.e., x_1, x_2, \dots, x_n are affected by noise, we will not obtain perfect energy compaction into one coefficient. However, we keep the subspace constraint \mathbf{c} as it reflects the vertex-weighted graph.

3. GRAPH-BASED ROTATION (GBR)

3.1. Rotation of Basis

Since the well-known discrete Fourier transform (DFT) diagonalizes circulant matrices, our goal is to find a rotation from the DFT to T such that T diagonalizes a covariance matrix, given the subspace constraint \mathbf{t}_1 .

Let L_0 be an $n \times n$ covariance matrix with circulant structure. An $n \times n$ unitary DFT matrix F is the matrix of eigenvectors of L_0 and computes the eigenvalue matrix Λ_0 . The first basis vector \mathbf{f}_1 of F is always a scaled all-one vector,

$$\mathbf{f}_1 = \frac{1}{\sqrt{n}} [1, 1, \dots, 1]^T. \quad (5)$$

Let us expand L_0 with the outer product of its eigenvector $\mathbf{f}_1 \mathbf{f}_1^T$, we have

$$L_0 = \alpha n \mathbf{f}_1 \mathbf{f}_1^T + (1 - \alpha) C, \quad (6)$$

where α is the weight of $n \mathbf{f}_1 \mathbf{f}_1^T$ and C is a complementary circulant matrix. Since C is also circulant, F again diagonalizes C . The eigenvalues of L_0 are then calculated as

$$\lambda_1 = \alpha n + (1 - \alpha) d_1, \quad (7)$$

$$\lambda_i = (1 - \alpha) d_i, \quad \text{for } i = 2, 3, \dots, n, \quad (8)$$

where d_1 and d_i are the diagonal entries of

$$D = F^T C F = \text{diag}[d_1, d_2, \dots, d_n]. \quad (9)$$

From the previous discussion, we want to preserve our subspace constraint $\mathbf{t}_1 = \frac{\mathbf{c}}{\|\mathbf{c}\|_2}$. We hope to find a new matrix L_1 such that it provides an eigenvector matrix T with the first basis vector \mathbf{t}_1 . To find a such T , we start with F . There exists a rotation matrix R that rotates \mathbf{f}_1 to \mathbf{t}_1 , i.e.,

$$\mathbf{t}_1 = R \mathbf{f}_1. \quad (10)$$

Thus, the rotation is determined by the first basis vectors \mathbf{f}_1 and \mathbf{t}_1 . Since we want to keep the same rotation for higher dimensional basis vectors \mathbf{f}_i and \mathbf{t}_i , $i > 1$, we have

$$T = R F. \quad (11)$$

Let Λ_1 be the corresponding eigenvalue matrix of L_1 . As the energy is preserved after the rotation, the sum of the eigenvalues for L_0 and L_1 are equal. Thus, for simplicity, we choose $\Lambda_1 = \Lambda_0$, i.e., L_1 gives the same eigenvalues as L_0 . Given the rotation R , L_1 can be calculated as

$$L_1 = R L_0 R^T. \quad (12)$$

Fig. 1 depicts a commutative diagram for L_0, L_1, F , and T . The problem of rotating F to T is related to the mapping from L_0 to L_1 . The link is given by the eigenvalue decomposition / Karhunen Loeve Transform (KLT).

From (12), we can write L_1 in a similar form to (6),

$$L_1 = \alpha n \mathbf{t}_1 \mathbf{t}_1^T + (1 - \alpha) T D T^T, \quad (13)$$

where D is given in (9). The unknown part in (13) is T , which indicates that if there exists a T that satisfies the subspace constraint \mathbf{t}_1 , L_1 can be obtained by (13). This L_1 in return gives T as an eigenvector matrix of L_1 with the first basis vector \mathbf{t}_1 . T always exists, as it can be found by, e.g., the Gram-Schmidt orthonormalization. However, T is not unique in high dimensions. Since neither L_1 nor R is known, T cannot be solved directly from the commutative diagram.

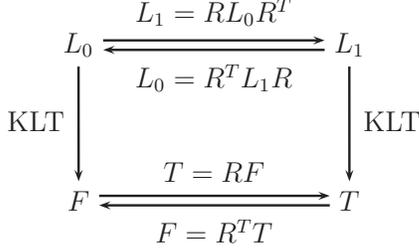


Fig. 1. Commutative diagram for covariance matrices and transforms.

3.2. Practical Algorithm

Since the transform matrix T is not unique in high dimensions, we choose a simple way to rotate the DFT matrix. To find a unique T , we rotate F only in the plane spanned by \mathbf{f}_1 and \mathbf{t}_1 , i.e., the projection of F onto this plane will be rotated only. With that, \mathbf{f}_1 will be rotated to \mathbf{t}_1 to meet the subspace constraint. The orthogonal part of F is left untouched. Let T' be a one-dimension-reduced matrix in T , i.e., $T = [\mathbf{t}_1, T']$. Similarly, let F' be a one-dimension-reduced matrix in F , i.e., $F = [\mathbf{f}_1, F']$. F' has the same dimension as T' .

To rotate F to T , we follow the steps below:

1. In the plane spanned by \mathbf{f}_1 and \mathbf{t}_1 ($plane(\mathbf{f}_1, \mathbf{t}_1)$), we find two vectors \mathbf{e}_1 and \mathbf{b}_1 , where $\mathbf{e}_1 \perp \mathbf{f}_1$ and $\mathbf{b}_1 \perp \mathbf{t}_1$. The projection of F' onto $plane(\mathbf{f}_1, \mathbf{t}_1)$ is the same as the projection of F' onto \mathbf{e}_1 , denoted as $proj(F', \mathbf{e}_1) \cdot \mathbf{e}_1$. And the projection of T' onto $plane(\mathbf{f}_1, \mathbf{t}_1)$ is the same as the projection of T' onto \mathbf{b}_1 , denoted as $proj(T', \mathbf{b}_1) \cdot \mathbf{b}_1$. Since the rotation is only performed in this plane, the projection weight does not change, i.e., $proj(T', \mathbf{b}_1) = proj(F', \mathbf{e}_1)$. Only the projected vector direction is changed from \mathbf{e}_1 to \mathbf{b}_1 .
2. The orthogonal part $O_{F'} \perp plane(\mathbf{f}_1, \mathbf{t}_1)$ in F' is calculated as

$$O_{F'} = F' - proj(F', \mathbf{e}_1) \cdot \mathbf{e}_1. \quad (14)$$

Similarly, the orthogonal part in T' is calculated as

$$O_{T'} = T' - proj(T', \mathbf{b}_1) \cdot \mathbf{b}_1. \quad (15)$$

Since both $O_{F'}$ and $O_{T'}$ are orthogonal to $plane(\mathbf{f}_1, \mathbf{t}_1)$, they are independent of the rotation. Thus, for simplicity, we keep $O_{T'} = O_{F'}$.

3. Based on (15), the matrix T' is then

$$T' = O_{F'} + proj(F', \mathbf{e}_1) \cdot \mathbf{b}_1. \quad (16)$$

Thus, T' is calculated by rotating the projection part $proj(F', \mathbf{e}_1)$ from \mathbf{e}_1 to \mathbf{b}_1 and adding back the orthogonal part $O_{F'}$.

T is obtained by simply putting \mathbf{t}_1 and T' together. At this point, there is no need to calculate the rotation matrix R itself, since we can construct T from F directly by projection and addition. Since $O_{F'} \perp \mathbf{t}_1$ and $\mathbf{b}_1 \perp \mathbf{t}_1$, we have $T' \perp \mathbf{t}_1$. As F is orthonormal and the projection weight does not change, T is also orthonormal. If the weights in a graph are all equal, then $\mathbf{t}_1 = \mathbf{f}_1$, there is no need to rotate \mathbf{f}_1 to \mathbf{t}_1 . In this case our T is the same as F .

4. EXPERIMENTAL RESULTS

Since the DFT matrix is complex, we use the discrete cosine transform (DCT) matrix instead in our experiments. We evaluate the energy compaction for the QCIF sequences *Foreman*, *Bus*, *City* and *Mobile*. We compare the motion-compensated orthogonal transform (MCOT) [6], our graph-based rotation (GBR), the subspace-constrained transform (SCT) [9], and the KLT for given graphs. The basis vectors for MCOT and GBR are only graph dependent. However, the basis vectors \mathbf{t}_k , $k = 2, 3, \dots, n$, for the SCT and the basis vectors \mathbf{t}_k , $k = 1, 2, \dots, n$, for the KLT are signal dependent and require covariance estimation. For a practical coding scheme, they need to be signaled to the decoder. At this point, we look at the energy compaction only. Practical coding schemes will be considered in future work.

The covariance matrix is estimated from samples that can provide the same vertex-weighted graphs. The vertex-weighted graphs are considered to be the same if they are formed by the same motion vectors and have the same vertex weights. In the experiments, we use graphs that are supported by at least thirty samples and that contain less than nine connections for a reliable estimation. MCOT and GBR are not signal dependent. However, we use only these reliable samples for all the transforms to ensure a fair comparison.

The transforms are performed on a hierarchical decomposition of a group of pictures (GOP). In the experiments, we choose a GOP size of sixteen. Each GOP is divided into four graph sets, where each set contains four successive frames. Each set outputs one temporal lowband and three temporal highbands. A GOP of sixteen frames requires two decomposition levels. Thus, on the first decomposition level, we obtain four lowbands from the four sets. The second level decomposition operates on these four lowbands. Since our graph-based transform does not require a fixed graph structure, it is possible to choose a different GOP size, rearrange the number of graph sets, and alter the number of decomposition levels. The graphs are defined by 16×16 block motion with a search range of ± 32 .

Tables 1 to 4 show the relative energy in the temporal sub-band frames on the second decomposition level for *Foreman*, *Bus*, *City*, and *Mobile*, respectively. Note, GBR provides the same lowband energy as MCOT and SCT, since they all share the same basis vector \mathbf{t}_1 . The KLT is given as a reference for the energy compaction of the first basis vector. The first ba-

sis vector of GBR approximates that of the KLT well, since it compacts lowband energy close to that of the KLT. This confirms the relevance of the subspace constraint.

Tables 1 to 4 show that for higher dimensional basis vectors \mathbf{t}_k , $k > 1$, GBR closes the gap to the optimal SCT, given the constraint. The second largest energy component of GBR is greater than that of MCOT, but not as good as SCT due to the sacrifice of signal dependency. In Table 4, the smallest energy component for GBR is only half of that for MCOT, which implies that GBR has better energy compaction in high dimensions. Thus, GBR has the potential of achieving a better coding performance. Note, the basis vectors of GBR are uniquely determined by the underlying motion information and the rotation of the DCT basis. No extra information needs to be transmitted to recover the GBR basis. This is not the case for the SCT.

Table 1. Relative energy in the temporal subband frames on the second decomposition level for *Foreman*.

	Low	High 1	High 2	High 3
MCOT	98.73%	0.27%	0.60%	0.40%
GBR	98.73%	0.68%	0.29%	0.30%
SCT	98.73%	0.98%	0.22%	0.07%
KLT	98.74%	0.95%	0.24%	0.07%

Table 2. Relative energy in the temporal subband frames on the second decomposition level for *Bus*.

	Low	High 1	High 2	High 3
MCOT	90.34%	3.36%	4.46%	1.84%
GBR	90.34%	5.74%	2.20%	1.72%
SCT	90.34%	8.07%	1.26%	0.33%
KLT	92.78%	5.73%	1.19%	0.30%

Table 3. Relative energy in the temporal subband frames on the second decomposition level for *City*.

	Low	High 1	High 2	High 3
MCOT	90.52%	2.78%	4.22%	2.48%
GBR	90.52%	5.00%	2.24%	2.24%
SCT	90.52%	7.20%	1.77%	0.51%
KLT	91.52%	6.38%	1.62%	0.48%

5. CONCLUSION

This paper presents a motion-adaptive transform that is based on a vertex-weighted graph and the rotation of the DCT basis. The vertex-weighted graph determines the first basis vector of the linear transform and leads to a subspace constraint. Our GBR is constructed by rotating the DCT basis, where the rotation is uniquely determined by the first basis vector of the

Table 4. Relative energy in the temporal subband frames on the second decomposition level for *Mobile*.

	Low	High 1	High 2	High 3
MCOT	92.89%	1.34%	4.31%	1.46%
GBR	92.89%	5.01%	1.43%	0.67%
SCT	92.89%	5.49%	1.27%	0.35%
KLT	92.99%	5.45%	1.21%	0.35%

DCT and the subspace constraint. The GBR closes the gap to the optimal energy compaction of the SCT, given our subspace constraint. The basis of the GBR does not need to be transmitted to the decoder side since it is independent of the signal.

6. REFERENCES

- [1] E. Martínez-Enríquez, F. Díaz-de-María, and A. Ortega, "Video encoder based on lifting transforms on graphs," in *Proc. of the IEEE International Conference on Image Processing*, Sep. 2011, pp. 3509–3512.
- [2] Woo-Shik Kim, S. K. Narang, and A. Ortega, "Graph based transforms for depth video coding," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2012, pp. 813–816.
- [3] F. R. K. Chung, *Spectral Graph Theory*, CMBS Regional Conference Series in Mathematics, No. 92. American Mathematical Society, 1997.
- [4] David K. Hammond, Pierre Vandergheynst, and Remi Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.
- [5] B. Girod and S. Han, "Optimum update for motion-compensated lifting," *IEEE Signal Processing Letters*, vol. 12, no. 2, pp. 150–153, 2005.
- [6] M. Flierl and B. Girod, "A motion-compensated orthogonal transform with energy-concentration constraint," in *Proc. of the IEEE International Workshop on Multimedia Signal Processing*, Oct. 2006, pp. 391–394.
- [7] M. Flierl and B. Girod, "Half-pel accurate motion-compensated orthogonal video transforms," in *Proc. of the IEEE Data Compression Conference*, Mar. 2007, pp. 13–22.
- [8] D. Liu and M. Flierl, "Video coding with adaptive motion-compensated orthogonal transforms," in *Proc. of the Picture Coding Symposium*, May 2012, pp. 293–296.
- [9] D. Liu and M. Flierl, "Motion-adaptive transforms based on vertex-weighted graphs," in *Proc. of the IEEE Data Compression Conference*, Mar. 2013, pp. 181–190.