# GROUP THEORETICAL TRANSFORMS AND THEIR APPLICATIONS IN IMAGE CODING

*Reiner Lenz and Jonas Svanberg*

Department of Electrical Engineering
Linköping University
S-58183 Linköping
Sweden

## ABSTRACT

We use the representation theory of finite groups to simplify the Karhunen-Loève Transform (KLT) of systems with group theoretically defined symmetries. In this paper we focus on applications of the dihedral groups D(n) which consist of all isometries that map the n-sided regular polygon into itself. The group D(4) is of special importance for all problems on square grids. Connected to each group is a type of Fourier Transform. This transform block-diagonalizes all operators that commute with the group operations. As a result all correlation matrices of processes with group theoretically defined symmetries are block-diagonalized. This simplifies the computation of the KLT considerably.

For real world data the symmetry assumptions leading to the simplification of the KLT are never exactly fulfilled and the KLT based on the block-diagonal correlation matrix is only an approximation to the correct KLT. In the second part of the paper we compare several approximations to the KLT for a large data base consisting of blocks collected from a standard TV-channel. Finally we discuss some of the consequences for image coding applications.

## 1. GROUP THEORETICAL TRANSFORMS

It is a well-known fact that the Fourier Transform of a signal is invariant up to a constant complex factor with respect to time-shifts. This relation between the Fourier Transform and the time-shift operators has a group theoretical generalization in which linear operators have the structure of a group. For these operator groups it is then possible to construct a corresponding Fourier-like transform with many properties similar to the Fourier Transform.

In the following we will only consider groups with finitely many elements. Many of the results are also valid for larger groups but we will not need them in this paper (further information can be found in the vast literature on the subject, see for example: [1], [2] and [3]). The connection between operators and a group $G$ is given by means of a **representation**:

**Definition 1** *Assume that $V$ is an $N$-dimensional vector space and that for each $g$ in a group $G$ we have a linear operator $T(g)$ that operates on $V$.*

1. *We say that $T$ forms an $N$-dimensional* **representation** *of $G$ if*
$$T(g_1g_2) = T(g_1)T(g_2)$$
*for all $g_1, g_2 \in G$.*

2. *Fixing a basis in $V$ we can describe each linear operator $T(g)$ by a matrix. Thus we define: an $N$-dimensional* **matrix representation** *of $G$ is a mapping: $T$ from $G$ into the space of $N \times N$ complex matrices such that*
$$T(g_1g_2) = T(g_1)T(g_2)$$
*for all $g_1, g_2 \in G$.*

3. *A base change in $V$ (described by a matrix $B$) describes the same linear operator by two different matrices: $T_1(g)$ and $T_2(g) = BT_1(g)B^{-1}$. We say that two matrix representations are* **equivalent** *if there is a matrix $B$ such that*
$$T_2(g) = BT_1(g)B^{-1}$$
*for all $g$.*

Selecting a "good" basis in $V$ might simplify the form of the matrices considerably. We say that the representation is **reducible** if there is a matrix $P$ such that

$$PT(g_k)P^{-1} = \begin{pmatrix} A_k & B_k \\ 0 & C_k \end{pmatrix} \quad (1)$$



0-7803-2431-5/95 $4.00 © 1995 IEEE

where the matrices $A_k, B_k$ and $C_k$ have the same size for all $k$. If a representation it is not reducible then we say that it is **irreducible**. If we can further simplify the representations so that they become block-diagonal with irreducible components along the diagonal ($B_k = 0$) we say that they are **completely reducible.**

For finite groups the representation theory states that there are only finitely many different irreducible representations (with respect to equivalence) and that all representations are completely reducible. Moreover the irreducible representations are all finite dimensional. For finite groups the role of the complex exponential function for the group of shift operations is played by the irreducible representations.

Of central importance for our application is the definition of group theoretical symmetries:

**Definition 2** *Let $T$ be an $N$-dimensional representation of a finite group $G$ and $C$ be an operator $C$ that acts on the same space as the $T(g)$. Then we say that $C$ is a* **G-symmetric operator** *if for all $g \in G$:*

$$T(g)C = CT(g). \tag{2}$$

The property of G-symmetric operators that is of interest for us is the following: All G-symmetric matrices are block-diagonalized in the coordinate system in which the representation matrices are given by irreducible blocks along the diagonal. The block structure, i.e. the number and size of the blocks is only determined by the group $G$ and it is the same for all G-symmetric $C$. The coordinate transform described by the matrix $P$ that accomplishes the change of coordinate system can be seen as a special transform associated with the corresponding group symmetry.

Of special interest are the groups of shift operations and the so called dihedral groups $D(n)$, defined as group of isometries that leave the regular n-sided polygon invariant. $D(n)$ consists of $2n$ rotations and reflections. In our application we are mainly interested in images on square (and perhaps also on hexagonal) grids. The relevant groups are in these cases D(4) and D(6). The main properties of the representations of D(2n) are summarized in the next theorem (for a complete description see [4])

**Theorem 1** *Assume that $n$ is even and denote by $\rho$ the rotation with rotation angle $\frac{2\pi}{n}$ and by $\sigma$ the reflection on a symmetry axis of the polygon. Set furthermore $\omega = e^{2\pi i/n}$. Then we have:*

1. *The elements in $D(n)$ are of the form $\sigma^l \rho^k$ with $(l = 0, 1; k = 0, \dots n-1)$.*
2. *There are four one-dimensional irreducible representations:*

$$T_1^1(\sigma^l\rho^k) = 1 \qquad T_2^1(\sigma^l\rho^k) = (-1)^l$$

$$T_3^1(\sigma^l\rho^k) = (-1)^k \qquad T_4^1(\sigma^l\rho^k) = (-1)^{k+l}$$

3. *There are $\frac{n}{2} - 1$ two-dimensional irreducible representations:*

$$T_j^2(\rho^k) = \begin{pmatrix} \omega^{jk} & 0 \\ 0 & \omega^{-jk} \end{pmatrix} \; T_j^2(\sigma\rho^k) = \begin{pmatrix} 0 & \omega^{jk} \\ \omega^{-jk} & 0 \end{pmatrix}$$

*with $j = 1, \dots \frac{n}{2} - 1$.*

It is remarkable that only roots of unity appear in these representations. To get real valued transforms these roots of unity can be replaced by the corresponding cosine and sine terms. As a result the D(4) transform can be implemented using only additions and subtractions.

In an important example from image coding the vector space $V$ has dimension 64 and is given by all vectors describing gray value distributions on $8 \times 8$ square image blocks. The elements of the $D(4)$-group act on these blocks as permutations that rotate and reflect these image blocks. The G-symmetric operator is given by the correlation matrix computed from these blocks. If all rotations and reflections of the blocks appear equally often then the correlation matrix is D(4) symmetric and the D(4) transform block-diagonalizes $C$.

## 2. EVALUATION OF KLT APPROXIMATIONS

In a rate/distortion sense the optimal transform for high bit-rate image transform coding is the **Karhunen-Loève transform** (KLT). By assuming some kind of group symmetry the group theoretical transform could be used as a preprocessing step which would simplify the problem of computing the KLT. In such an application one would first apply the group theoretical transform to the original data. Then one would assume that the correlation matrix for the transformed data is block-diagonal. The KLT is then computed for the resulting (smaller) blocks and the eigenvectors obtained are finally transformed back into the original domain. Ignoring all entries outside the blocks on the diagonal leads to errors in the computed eigenvectors and the true KLT will only be obtained when the problem was fully G-symmetric.

In our experiments we evaluated a number of KLT approximations with the help of a large database. The database consisted of 100 000 patches of size $8 \times 8$ which were digitized from a Swedish television channel over a period of 10 days. We collected the patches by digitizing 10 000 frames with a commercial frame grabber. In each of the frames 10 randomly located blocks were processed. For each pixel we got a three-component vector with the RGB values at this location. From the RGB vectors we computed the $3 \times 3$ correlation matrix and its eigenvectors. The eigenvalue distribution showed that most of the information was contained in the first eigenvector which acts approximately as averaging filter. Using this first eigenvector we decorrelated the color information pixelwise and projected the RGB-vector down to a one-dimensional intensity value. These patches of intensity value distributions are the data vectors used in the following experiments.

In one series of experiments we tested five different KLT approximations against the true eigenvector system:

**D(4)-symmetry:** the second order statistics are unchanged under reflections and rotations by multiples of 90 degrees.

**Stationarity:** the autocorrelation function is only a function of the difference vector on the grid.

**Circular-symmetry:** correlations depend only on the distance between points.

**Toeplitz** the autocorrelation function can be separated into two one-dimensional autocorrelations.

**DCT** shift invariance leading to a fixed transformation as KLT approximation.

For each approximation method we computed first the approximated eigenvector system from all patches in the database. Then we selected 1000 random patches from the database and computed the reconstruction error for all truncation orders. By plotting the reconstruction errors for all approximation orders we can compare the performance of the different methods (see Figure 1). These experiments showed that the stationary assumption is clearly valid. We conclude also that the dihedral approximation fits the data reasonably well, whereas the Toeplitz and the radial approximations are of significantly lower quality. A more detailed description of this experiment can be found in [5]. We also applied the different transforms to widely used images such as Lenna. We found comparable reconstruction errors for approximations with 1 to 8 terms. If we selected more terms, however, (about 20) then the ranking of the different methods is more or less exactly the opposite of what could be expected from our previous experiments: Now the radial approximation works best and the Karhunen-Loéve expansion is among the worst. This shows that there are significant statistical differences between these images and the patches collected from the TV-channel.

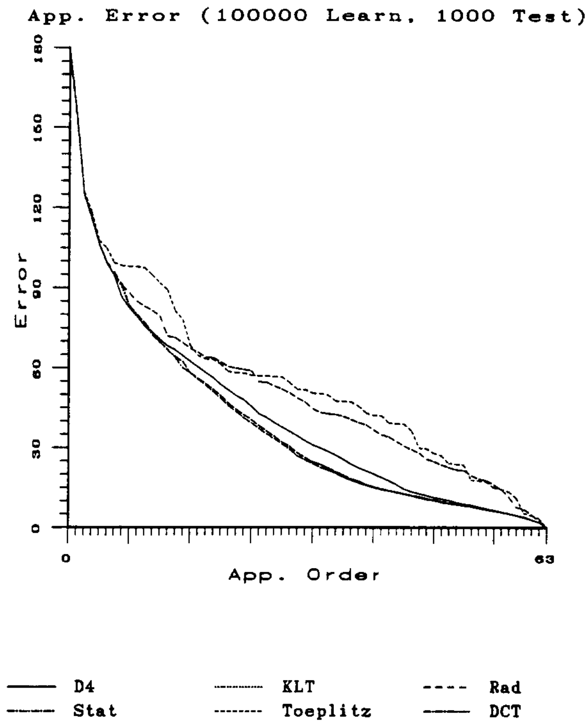

Figure 1: Reconstruction error for the $8 \times 8$ database

## 3. IMAGE CODING APPLICATIONS AND SUMMARY

One advantage of the $D(4)$-transform over the DCT is that it is not restricted to rectangular blocks. The selection of the form of the window is only restricted by the requirement that the window must be invariant under rotations and reflections. For all these blocks the $D(4)$-transform is applicable. It has been argued that such alternative tiling schemes might have some advantages in image coding since the eye seems to be more sensitive to horizontally and vertically oriented artifacts (some experiments with different tilings of the image are reported in [6]).

Some results of our preliminary experiments with such non-standard tiling methods is shown in Figures 2 and 3. In this experiment we partitioned the image with two different tiles: a snowflake-like and a more circular shaped tile. From the two correlation matrices belonging to these tiles we computed a D(4) approximation of the KLT. We then used these two different sets of basis vectors in our coding experiment. In

Figure 2 we see the coded versions of Lenna with the DCT (top) and the non-standard D(4) tiling (bottom) respectively. In diagram 3 we show the PSNR for a number of different compression rates. The diagram shows that the numerical performance of the DCT-based method is always better than the D(4) based coding. The question if this numerical advantage also corresponds to a perceptual advantage is harder to answer and has to be tested with a larger number of images.

Summarizing, we showed that for each finite group there is a DFT-like transform and that the application of this transform block-diagonalizes operators with group theoretical symmetries. In cases where the operators do not posses the full symmetry we get approximations to the KLT. We concentrated mainly on the description of the dihedral transforms and evaluated this and several other popular approximations to the KLT with the help of a large database of image blocks. The experiments showed that the symmetry assumptions are largely valid but they showed also that widely used standard images have different statistical characteristics than these TV-images. We also studied some image coding applications and compared the D(4) based methods with the standard DCT coding scheme.

## 4. REFERENCES


[1] Reiner Lenz, "Using representations of the dihedral groups in the design of early vision filters", in *Proc. International Conference on Accoustics, Speech and Signal Processing.* 1993, pp. V165–V168.

[2] Reiner Lenz, *Group Theoretical Methods in Image Processing*, Lecture Notes in Computer Science (Vol. 413). Springer Verlag, 1990.

[3] C. W. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras*, Wiley Classics Library. Wiley, New York, 1988.

[4] Reiner Lenz, "Group theoretical transforms in image processing", in *Current Topics in Pattern Recognition Research.*

[5] Reiner Lenz, "Statistical properties of TV-signals and approximations of the Karhunen-Loéve-transform", Tech. Rep., Dept. EE., Linköping University, S-58183 Linköping, Sweden, 1994.

[6] R. M. Gray, P. C. Cosman, and K. L. Oehler, "Incorporating visual factors into vector quantiziers for image compression", in *Digital Images and Human Vision*, Andrew B. Watson, Ed., chapter 4, pp. 35–52. MIT PRESS, Cambridge, Mass., London, England, 1993.


Figure 2: Coding with DCT (top) and D(4) Tiling(bottom)

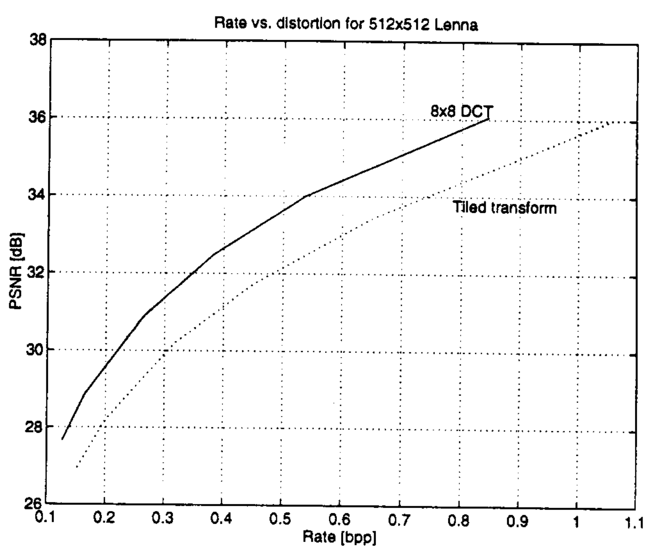


Figure 3: Comparison of the PSNR for DCT and D(4) coding