The Domain of Orthogonal Transforms and the Understanding of Image Features

Malek Adjouadi, Habibie Sumargo, Jean Andrian, and Frank Candocia

Department of Electrical and Computer Engineering
Florida International University
University Campus, Miami, Florida 33199, USA

Abstract

In this study, we explore the domain of orthogonal transforms, in order to bring an understanding on the characterization of image features, with emphasis placed on the Karhunen-Loeve (K-L) transform for its optimal energy packing properties. The study's contribution is in establishing a thorough research base that relates the eigensystem and transform domain properties of the K-L transform to two-dimensional image features. Other transformations such as Haar, Hadamard, Walsh, Fourier, and the discrete cosine transform (DCT) are presented as a basis for performance evaluation in the energy packing sense. The main thrust of this study, given the transform domain, is to determine mechanisms and develop algorithms relevant to the understanding and discrimination of visual features invariant across translation, orientation, gray-level polarity reversal, and size. The computational requirements are addressed.

Introduction

The Karhunen-Loeve (K-L) transform is recognized as the optimum transformation in the sense that it packs the most energy into its first few transform coefficients as it minimizes the mean square error between the original image and its corresponding reconstructed image from a smaller set of uncorrelated transformed data. This is an important feature for data compression. Also, K-L transform yields features most desirable for the development of algorithms that are invariant across orientation and translation. However, it has always been dismissed for its computational complexity due to the fact that the eigensystem need to be evaluated requiring the computation of large inverse transformations. This is compounded by the fact that no fast algorithm can be implemented for such a transformation. For this reason, an implementation of the Household method to tridiagonalize the covariance matrix in order to expedite the finding of the eigensystem is proposed. As it will be made clear, the drawback of computational complexity should be reassessed for all the advantages provided by the K-L transform.

The focus is placed on the understanding of the relational principles that exist between image features and their subsequent change due to rotation, translation and other variations with respect to the noted effect resulting in the transform domain. Some of the visual features understanding algorithms developed in this study draw from the work by Hubel and Wiesel on "the visual cortex and the seeing of features" [1,2]. An analysis is provided relating the results obtained on the transform domain to image features in context with the effect of edge lines with different orientations, positions, and other morphological properties. The preliminary results obtained for this particular problem constitute but a simple first step of what may be a fundamental problem in vision, that of bringing some credence to the belief that transformations may be involved in the important issues of size constancy and orientation-independence.

A unified mathematical framework is provided to relate in a simple way all of the orthogonal transforms. It is then easier to perform comparative studies, and characterize uniquely each one of the transform in their energy packing sense, as well as in their responses to image features and properties. Certain mathematical fundamentals such as the eigensystem, the covariance matrix, the correlation matrix, and other statistical parameters are exploited for optimized visual features interpretation and understanding. Throughout this study, all computer implementations are assessed by means of their performance in relation to the computational requirement, and in their effectiveness to visual features understanding.

Orthogonal Transforms Mathematical Framework

The orthogonality principle of transforms and the separability of the transform kernels are best understood considering the following mathematical framework:

A matrix M as described below:

$$M=\begin{bmatrix} v_{1_c} & v_{2_c} & \cdots & v_{k_c} \end{bmatrix}=\begin{bmatrix} v_{1_r} & v_{2_r} & \cdots & v_{k_r} \end{bmatrix}^T,$$

is said to be unitary if $\left(M^T\right)^*=M^{-1}$, where $\left(M^T\right)^*$ is the conjugate transpose of M. If M is a real matrix, then M is an orthogonal matrix such that $M^T=M^{-1}$. These type of orthogonal matrices satisfy the following conditions:

$$v_{i_c}^T\cdot v_{j_c}=\begin{cases}0 & \text{if } i\neq j\\ cte & \text{if } i=j\end{cases}, \text{ and } v_{i_r}\cdot v_{j_r}^T=\begin{cases}0 & \text{if } i\neq j\\ cte & \text{if } i=j\end{cases}$$

The transformations used in image processing are in general orthogonal transformations which yield transform elements that are highly decorrelated. The orthogonality property is what allows us to avoid the burden of computing the inverse transformation.

As an illustration, recall that the discrete Fourier transform (DFT), can be written as:

$$F(u,v)=\frac{1}{N}\cdot\sum_{x=0}^{N-1}e^{\frac{-j\cdot 2\cdot\pi\cdot(u\cdot x)}{N}}\cdot\left[\sum_{y=0}^{N-1}f(x,y)\cdot e^{\frac{-j\cdot 2\cdot\pi\cdot(v\cdot y)}{N}}\right]$$

where $u,v=0,1,\cdots,N-1$

thus, F(u,v) can be written in matrix form as:

$$F(u,v)=k_f(x,u)\cdot f(x,y)k_f(y,v)^{*T}$$

where $k_f(y,v)^{*T}$ denotes the transpose conjugate of the forward kernel, $k_f(y,v)$. Such is the case of complex transformations like the Fourier transform. For real orthogonal transforms, the general formulation for the transformation operation can be written as:

$$T(u,v)=k_f(x,u)\cdot f(x,y)\cdot k_f(y,v)^T$$

Using the same principles, the inverse transformation operation can be written as:

$$f(x,y)=k_i(x,u)\cdot T(u,v)\cdot k_i(y,v)^T$$

where $k_i=k_f^{-1}=k_f^T$; $f(x, y)$ can be written as

$$f(x,y)=k_f^T(x,u)\cdot T(u,v)\cdot k_f(y,v)$$

note that $k_f(y,v)$ and $k_f(x,u)$ are functionally equal.

2-D Image and Relevant Statistical Descriptors

If we consider a two-dimensional (2-D) image X, then the covariance matrix of X is defined as:

$$\sum\nolimits_x=\begin{bmatrix}\sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2\\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2n}^2\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_{nn}^2\end{bmatrix}$$

where the variance elements are given by:

$$\sigma_{ij}^2=\frac{1}{n}\cdot\left[(row_i)\cdot(row_j)^T\right]-(M_{row\,i})\cdot(M_{row\,i})^T$$

and the mean values are given by:

$$M_{row\,i}=\frac{1}{n}\cdot\sum_{j=0}^{n-1}x_{ij}$$

x_{ij} being the elements of X. The correlation matrix is defined as:

$$R=\begin{bmatrix}1 & r_{12} & \cdots & r_{1n}\\ r_{21} & 1 & \cdots & r_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ r_{n1} & r_{n2} & \cdots & 1\end{bmatrix},\qquad |r_{ij}|\le 1$$

where,

$$r_{ij}=\begin{cases}\dfrac{\sigma_{ij}^2}{\sigma_{ii}\cdot\sigma_{jj}}, & i\ne j\\ 1, & i=j\end{cases}$$

An elaborate study of these statistical descriptors is given in [3-5].

It is noted that the covariance matrix is real and symmetric. It is the symmetry property that is exploited in using the Householder method, as will be shown later, in order to expedite the finding of the eigensystem.

Application of the Karhunen-Loeve Transform (KLT)

Like other transforms, the K-L is an orthogonal transform but statistical in nature. That is, it performs a statistical analysis on the variation in the image data. Considering an n x n input image X, the KL transform can be derived as:

$$y=A\cdot(X-m_X)$$

where A is the matrix composed of eigenvectors of the covariance matrix of X, and m_x denotes the mean vector of X found by averaging the values of each row. It can be shown that the covariance matrix of y can be computed as:

$$C_y=A\cdot C_X\cdot A^{-1}$$

C_y is a matrix whose elements are zero except along the main diagonal where the values are the eigenvalues of C_X. Since C_y and C_X have the same eigenvalues, C_y and C_X will have the same eigenvectors. Using the orthogonal property of the eigenvectors, $A^{-1}=A^T$, it is possible to reconstruct input image from the transform as follows:

$$\hat{X}=A^T\cdot y+m_X$$

This inverse transformation is a one-to-one mapping (lossless transformation) if we preserve all the eigenvectors. However, for data compression purposes, if we take into account only the energy from the first k eigenvectors that correspond to the first k highest eigenvalues, then the KLT can be expressed as follows:

$$y_k=A_k\cdot(X-m_X)$$

where A_k is composed of the first k eigenvectors and both A_k and y_k are of a k x n dimension. The lossy reconstruction is thus:

$$\hat{X}_k=A_k^T\cdot y_k+m_X$$

where A_k^T and $\hat{X}_k$ are of n x n dimension.

Application of the Householder Method

The Householder method reduces an n x n real and symmetric matrix to a tridiagonal matrix using (n-2) Hermitian ($X^{T*}=X$) transformations [6]. Since input images are not necessarily symmetric, a real and symmetric covariance matrix of the input image is first determined before the Householder method is used. The iterative process of the Householder method annihilates the required part of a whole column and whole corresponding row for each iteration. Consider an n x n real and symmetric matrix X such that:

$$X_{r+1}=T^{(r)}\cdot X_r\cdot T^{(r)},\quad \text{where } r=1,2,\cdots,n-2$$

The $T^{(r)}$ matrix is the Householder matrix defined as:

$$T^{(r)} = I - 2 \cdot W^{(r)} \cdot W^{(r)^T}, \text{ where}$$

$$W^{(r)} = \alpha \cdot \{0_1, \cdots, 0_r, V_{r+1}^{(r)}, y^T\}, \text{ where}$$

$$\alpha = \frac{1}{\sqrt{2} \cdot \left(\sqrt{S^2 + \operatorname{sign}(x_{r+1}) \cdot S}\right)}, \text{ where } S^2 = x_{r+1}^2 + (y^T \cdot y)$$

the variables $0_1, \cdots, 0_r$ denote the inserted zeroes at the r^{th} iteration. The vector y consists of the element from $(r+1)$ to n of row r, and

$$V_{r+1}^{(r)} = x_{r+1} + \operatorname{sign}(x) \cdot S$$

x_r is the r^{th} row of matrix X. The whole process of tridiagonalizing an n x n matrix through Householder method requires H_n operations [7], where H_n is given as follows:

$$H_n = \frac{1}{3} \cdot (n-2) \cdot (2 \cdot n^2 + 7 \cdot n + 27)$$

Once tridiagonalization is achieved, the eigensystem can be computed more efficiently through the use of QL algorithm with implicit shift given in [8].

Computational Aspects

Below is a timing comparison of the computational requirements of an n x n matrix between standard evaluation of the determinant and using the Householder method.

a. Standard Method

The determinant of a 2 x 2 matrix is:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$$

This requires 2 (mults) and 1 (add/sub) operations that is,

$$O_2 = \varepsilon_0 = 2 \cdot (M) + 1 \cdot (A)$$

Similarly, for a 3 x 3 matrix, we find:

$$a_{11} \cdot cof[a_{11}] - a_{12} \cdot cof[a_{12}] + a_{13} \cdot cof[a_{13}]$$

where $cof[a_{ij}]$ is the co-factor of a_{ij} element. This co-factor requires O_2 operations. Thus, the number of operations required to compute the determinant of a 3 x 3 matrix is:

$$O_3 = 3 \cdot (\varepsilon_0) + 3 \cdot (M) + 2 \cdot (A)$$

For a 4 x 4 matrix, then:

$$O_4 = 4 \cdot [3 \cdot (\varepsilon_0) + 3 \cdot (M) + 2 \cdot (A)] + 4 \cdot (M) + 3 \cdot (A)$$

operations are required. The number of operations required to compute determinant of a 5 x 5 matrix is:

$$O_5 = 5 \cdot [4 \cdot [3 \cdot (\varepsilon_0) + 3 \cdot (M) + 2 \cdot (A)] + 4 \cdot (M) + 3 \cdot (A)] + 5 \cdot (M) + 4 \cdot (A)$$

As we can see that if the order of the matrix increases, so does the number of operations. For an n x n matrix, the number of operations required is in the order of $n!$.

b. Householder Method

After tridiagonalization, the evaluation of the determinant of a 3 x 3 Householder matrix is as follows:

$$a_{11} \cdot cof[a_{11}] - a_{12} \cdot cof[a_{12}]$$

This requires:

$$O_3 = 2 \cdot (O_2) + 2 \cdot (M) + 1 \cdot (A) = 2 \cdot (\varepsilon_0) + \varepsilon_0 = 3 \cdot \varepsilon_0$$

Similarly, for a 4 x 4 matrix, it is determined that:

$$O_4 = 2 \cdot [3 \cdot (\varepsilon_0)] + \varepsilon_0 = 5 \cdot \varepsilon_0$$

Thus, by induction, for an n x n matrix, the number of operations required to compute the determinant of a tridiagonalized matrix is:

$$O_n = (2^n - 1) \cdot \varepsilon_0$$

Below is a table on the timing analysis in computing the determinant of an n x n matrix: H_n denotes the amount of computations required for the tridiagonalization process.

Size	H_n	$\det(H_n)$	$H_n + \det(H_n)$	Standard det
3	22	9	31	6
4	58	21	79	24
8	422	381	803	40320
16	3038	196605	199643	$2.1 * 10^{13}$

Image Features and the Transform Domain

Energy Conservation

In this section, a comparison of several orthogonal transforms is given in function of the energy distribution. The transforms given are: Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT), Haar Transform (HaT), Hadamard Transform (HT), Walsh Transform (WT), and Karhunen-Loeve Transform (KLT). Figure 1 illustrates the ways this energy is packed by the different orthogonal transforms. Figures 1(a) and 1(b) illustrate the case where variances of rows and columns of the given image are all the same ($\sigma_{ij}^2 = c$, where i=1,..,k, and j=1,..,k), and the case where an abrupt change exists between any two pixels, respectively. The transformations used in Figure 1 take the form:

$$[T(u,v)] = \sigma^2 \cdot [k_f \cdot R_X \cdot k_f']$$

noting that the covariance matrix $\Sigma_X = \sigma^2 \cdot R_X$, where k_f is the forward transform kernel.

Observation: The energy of a scene containing two objects is equivalent to the sum of the energies of the two objects.

	Eigenvalues								Energy
	1	2	3	4	5	6	7	8	E_t
Object 1	38368	4305	0	0	0	0	0	0	42673
Object 2	36378	11097	3326	0	0	0	0	0	50801
Object 1+2	54888	20589	11041	4328	2628	0	0	0	93474

where E_t denotes the total energy of the transforms obtained from the eigenvalues.

Effect of Orientation

Observation 1: Regardless of the orientation of an object, if the contour of the object is not affected by the orientation, the K-L transform yields the same eigensystem and the same energy.

Observation 2: Any amount of object information hidden due to the effect of orientation is reflected proportionally in K-L transform domain.

Below is a table reflecting a constant eigensystem and energy conservation for an object that is rotated at different angles.

	1	2	3	4	5	6	7	8	E_t
Scene-1	13243	3441.7	1633	1008.1	729.1	590.5	524	0	21169.4
Scene-1, rotated 90°	13243	3441.7	1633	1008.1	729.1	590.5	524	0	21169.4
Scene-1, rotated 180°	13243	3441.7	1633	1008.1	729.1	590.5	524	0	21169.4
Scene-1, rotated -90°	13243	3441.7	1633	1008.1	729.1	590.5	524	0	21169.4

where E_t denotes the total energy of the transforms obtained from the eigenvalues.

Size Constancy

Observation: It is noted that a change in size of an object is reflected proportionally in the energy of the K-L transform domain.

	Eigenvalues								Energy
Object size	1	2	3	4	5	6	7	8	E_t
S_1	485350	34850	0	0	0	0	0	0	520200
S_2=2 S_1	864280	176120	0	0	0	0	0	0	1040400
S_3=3 S_1	1211600	349000	0	0	0	0	0	0	1560600

where E_t denotes the total energy of the transforms obtained from the eigenvalues.

Effect of Translation:

Observation 1: Translation of an object on a homogeneous background has no effect on the K-L transform domain. The eigensystem and the energy remain constant.

Observation 2: An image containing an object in translation on a homogeneous background with respect to a stationary object, would yield a transform domain where the eigensystem may be different but with the same amount of energy, unless there is an overlap between the objects. The difference in the eigensystem is function of the way the covariance matrix is computed (row wise vs. column wise), and the placement of these objects in the scene.

	Eigenvalues								Energy
	1	2	3	4	5	6	7	8	E_t
Object 1	38368	4305	0	0	0	0	0	0	0
Object 1-T	38368	4305	0	0	0	0	0	0	0

where E_t denotes the total energy of the transforms obtained from the eigenvalues.

Reversal in Gray-Level Polarity

Observation: A change from a positive image to a negative image of the same scene yields no change in the eigensystem as well as in the energy.

Conclusion

The results presented in this study were obtained using various synthetic and real images. The Householder method improved significantly the processing time of the K-L transform from the order of $n!$ computations to the order of $(2^n - 1)$ computations in finding the eigensystem of large matrices. Various aspect of the image features were accessed in terms of the eigensystem and other statistical descriptors. Such type of analysis leads effectively to the development of algorithms that are invariant across translation, rotation, and reversal of gray-level polarity.

References

[1] J. P. Frisby, Seeing: Illusion, Brain and Mind, Oxford University Press, 1978.

[2] D. H. Hubel, Eye, Brain, and Vision, W. H. Freeman and Co., New York, 1988.

[3] W. K Pratt, Digital Image Processing, John Wiley & Sons, Inc., New York, 1991.

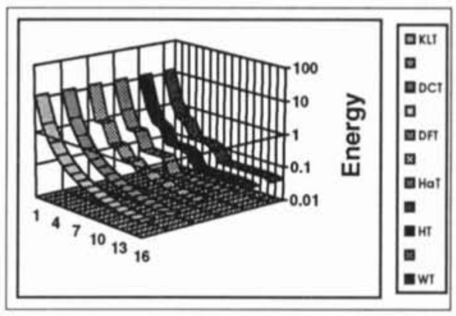
[4] R. C. Gonzalez and R. E. Woods, Digital Image Processing, Addison-Wesley Publishing Company, Reading, Massachusetts, 1992.

[5] N. Ahmed and K. R. Rao, Orthogonal Transformation in Digital Image Processing, Springer-Verlag, New York, 1975.

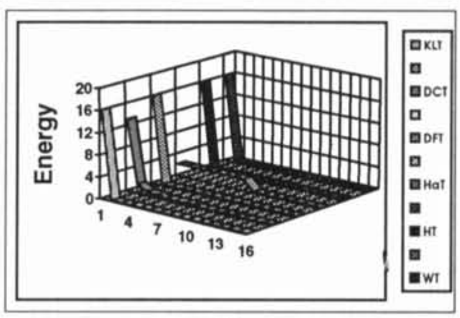
[6] A. S. Householder, The Theory of Matrices in Numerical Analysis, Blaisdell Publishing Company, New York, 1964.

[7] S. J. Hammarling, Latent Roots and Latent Vectors, The University of Toronto Press, Toronto, 1970.

[8] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Numerical Recipes in C, Cambridge University Press, New York, 1988.



(a)



(b)

Figure 1. Energy packing feature of transforms