LINEAR TRANSFORMATION IS BAD FOR THE CODING OF GRAPHIC IMAGES

S.T. Vuong

Department of Computer Science
University of Waterloo

ABSTRACT

Although the coding technique using linear transformations has proven practical for continuous tone images, its applicability to graphic images has not been examined in the literature. In this paper the feasibility of employing the transform coding technique in the compression of binary pictorial data is investigated. Typical linear transformations are reviewed and a hybrid tree-transform coding method is introduced, which promises to perform better than the simple transform scheme for graphic images. Simulation results from the processing of five different graphic image samples on the Image-100 system strongly suggests, however, the impracticability of using linear transformations in the coding of graphic images.

RÉSUMÉ

Même si la technique de codage qui utilise les transformations linéaires s'est avérée pratique pour des images à tonalité continue, son applicabilité aux images graphiques n'a pas été étudiée. Dans le présent document, on analyse la faisabilité de l'utilisation de la technique de codage transformée pour la compression des données picturales binaires.

Les transformations linéaires types sont étudiées et une méthode de codage en arborescence hybride est introduite, méthode qui promet de donner de meilleurs résultats que la simple transformation de schémas en images graphiques. Les résultats de simulation obtenus à partir du traitement de cinq échantillons d'images graphiques différentes sur le système Image-100 laisse toutefois croire fortement à l'impossibilité d'utiliser les transformations linéaires dans le codage des images graphiques.

I. INTRODUCTION

Although the transform coding technique has proven practical for continuous tone images [1], its applicability to graphic images, to my knowledge, has not been investigated in any work in the literature. The basic idea of transform coding is to map the sets of correlated picture elements (pixels) into the sets of nearly uncorrelated coefficients which can be encoded efficiently. For graphic images, however, there seems to be a contradiction. On the one hand, the strong correlation among neighboring picture elements in graphic images suggests the tractability of the transform coding method in the compression of binary pictorial data. On the other hand, as linear transformations generally map binary data into multilevel data, it appears unwise to go to the transform domain to code graphic images. In this paper, the feasibility of employing the transform coding technique in the compression of binary pictorial data is investigated. Typical linear transformations and transform coding schemes are first reviewed. A hybrid tree-transform coding method is then introduced, which promises to perform better than the simple transform scheme for graphic images. Finally, a computer experiment performed on five picture samples is described and simulation results are discussed.

II. LINEAR TRANSFORMATIONS AND TRANSFORM CODING

The basic premise of the transform coding system is that the transform of an image has an energy distribution more suitable to coding than the representations in the spatial domain. Figure 1 shows the block diagram of a transform coding system. Exploring both statistical and psychovisual redundancies in images, the transform encoder performs basically a sequence of two operations. The first operation is a linear transformation, A, which transforms a set of statistically dependent picture elements (pixels) f into a set of "more independent" coefficients, F. The second operation, Q, is to quantize and code each coefficient. The number of bits required to code each coefficient depends on the number of quantizer levels dictated by the sensitivity of human vision to subjective effect of quantization error.

The best transformation for image coding would be the one resulting in a set of statistically independent coefficients. Needless to say, it is practically impossible to obtain this transformation, and the closest one can get to such an ideal with a linear transformation is the one that produces uncorrelated coefficients such as the discrete Karhunen-Loeve. Before examining the applicability of the transform coding technique in the compression of binary pictiorial data, it is worthwhile to review typical linear transformations and transform coding schemes that have been proposed for the coding of multilevel images.

The discrete Karhunen-Loeve transformation.

For the sake of simplicity, let us consider the one-dimensional transformation:

$$F = Af$$

Where f and F denote the image vector and the image transform vector respectively. Then, The discrete Karhunen-Loeve transformation matrix is the modal matrix M of the covariance matrix R_f of the image vector. We can write:

$$A = M^T = [\xi_1 \, \xi_2 \ldots \xi_N]^T$$

with ξ_i denoting the eigenvector corresponding to the i^{th} largest eigenvalue of R_f. It can be shown that this transformation results in uncorrelated coefficients. Indeed, the covariance matrix R_F of the coefficients can be expressed as:

$$R_F = E[FF^T] = (F.F) = (Af.Af) = A\,(f.f)\,A^T = A\; R_f\, A^T$$

where E[.] denotes the expected value and (..), the inner product . Since $A = M^T = M^{-1}$ we have

$$R_F = M^{-1} R_f M = \Lambda$$

Hence, R_F is a diagonal matrix with diagonal elements being eigenvalues of R_f. Consequently, the components of F are uncorrelated.

To achieve data compression, we can retain only the first m coefficients having the largest variances, i.e. the largest eigenvalues of R_f for minimun mean square error.

Although the discrete Karhunen-Loeve transformation appears theoretically attractive as it minimizes the mean square error, there are two major problems associated with its use [2]: i) much computation must be performed: the covariance matrix R_f must be estimated if not known; next it must be diagonalized to compute its eigenvalues and eigenvectors; then the transform itself must be taken. In general there is no fast computational algorithm for the transform. And, ii) mean square error is not a valid error criterion for many types of images.

Owing to the computational simplicity, a unitary transformation ($A^{-1} = A^{*T}$) such as the Fourier, Hadamard or Haar has been found or to be practical for image coding ([2], [3], and [4]).

The Discrete Fourier Transformation

The discrete Fourier transform pair can be expressed as

$$F = \frac{1}{\sqrt{N}} A f \text{ and } f = \frac{1}{\sqrt{N}} A^{*} f$$

where A^{*} denotes the conjugate of A which is a unitary and symmetric matrix with elements given by:

$$a_{kl} = W^{(k-1)(l-1)} \text{ and } W = e^{-2\pi i/N} \text{ for } k,l = 1,2,\ldots,N$$

If the image vector f has N real, positive components then its transform F has N complex components but as conjugate pairs. For $N = 2^m$, the fast computational algorithm requires (N/2)logN complex multiplications and NlogN complex additions and subtractions.

The Hadamard (or Walsh) Transformation

The Hadamard transform pair can be defined by:

$$F = H f \text{ and } f = (1/N) H F$$

where the Hadamard matrix H is an orthogonal and symmetric matrix with elements being 1's and -1's. For N=2, the Hadamard matrix H is defined as:

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

For N = 2^n, H can be constructed recursively by:

$$H_N = H_{N/2} \times H_2 = \begin{bmatrix} H_{N/2} & H_{N/2} \\ H_{N/2} & -H_{N/2} \end{bmatrix}$$

where $\times$ denotes the Kronecker product. A Walsh matrix is simply a Hadamard matrix with its basis vectors rearranged in the order of ascending sequency (i.e. the number of sign changes in a basis vector).

The implementation of a Hadamard (or Walsh) transform is extremely simple, and the fast computational algorithm using matrix factorization requires only NlogN additions or substractions.

The Haar Transformation

The Haar transform pair can be written as:

$$F = H f \text{ and } f = \frac{1}{N} H^T F$$

where the Haar matrix H is an orthogonal, non-symmetric matrix comprised of 1's, -1's and 0's and directly related to Walsh transform [5]. In term of sampling theory Haar matrix samples the input signal of progressively coarser intervals, starting with highest resolution and decreasing in powers of two. For example,

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix}$$

The Lower Triangular Transform

The lower triangular transformation is a non-orthogonal transformation which results in uncorrelated coefficients. In series form, the lower triangular can be written as:

$$F(1) = f(1)$$

$$F(u) = f(u) - \sum_{x=1}^{u-1} t_{ux} f(x)$$

$$\text{for } u = 2,3,\ldots,N$$

or in vector form,

$$F = Tf$$

where T is a unit lower triangular matrix such that the covariant matrix of F, $R_F = TR_f T^T$, is diagonal. Martin and Wilkinson [6] have developed an efficient algorithm for finding T and R_F reqriring only $N^3/6$ multiplications.

For an n^{th} order Markov process, the transform is given by:

$$F(u) = f(u) - \sum_{x=1}^{n} \alpha_x f(u-x)$$

where f(x) = 0 for x = 0, -1, -2,.... Then the operator T_n is a banded matrix of (n+1) bands. The transformation with operator T_n reqrires less than nN multiplications as compared to { $\frac{N^2}{2}$ - N} multiplications needed by the transformation with unit lower triangular operator in its general form.

Other interesting linear transformations with fast transform algorithms that are not described here include the cosine and the slant transformations.

Transform Coding Schemes

After the pixels of a picture (or subpicture) have been transformed into a set of coefficients, a coding scheme such as zonal coding or threshold coding can be applied to those coefficients to achieve data compression. In the zonal coding system, the set of zones is established in each transform block. Transform samples in each zone are then quantized with the same number of quantization levels which is normally set proportional to the expected variance of the transform coefficients. Thus, zero quantization level implies that the coefficients falling into that zone are discarded. In threshold coding, only those transform samples whose magnitudes are greater than a given threshold are retained and quantized with a fixed number of levels. Therefore, it is necessary to code not only the magnitude of each retained sample but also its position in the transform plane. Due to its adaptive process, the threshold coding system is expected to perform better than the zonal coding one. However, the implementation of the former would be much more complex than the one of the latter. On this account, the zonal coding strategy is normally used.

In general, a picture to be coded is partitioned into small equal size subpictures and a transform scheme is applied to each subpicture rather than the entire picture directly to reduce the computation and storage requirement.

III. TRANSFORM CODING FOR GRAPHIC IMAGES

As mentioned previously, the transform coding system has proven efficient and practical for continuous tone imagery. For binary pictorial data, however, the transform technique seems to be untractable despite the strong correlation among neighboring pixels. As a linear transformation generally maps binary data to multilevel data, many bits would be required to code each retained transform coefficient. Let us consider, for example, an one-dimensional Hadamard transform scheme with subpicture size N=16. Each subpicture of 16 binary pixels is then transformed into 16 coefficients, each having an integer value in the interval [-16,+16], i.e. the Hadamard transform maps binary data into data of 32 levels. Now, even a "crude", efficient coding scheme such as the one which uses 3 bits to code the DC coefficient, one bit to code each of the next 5 largest variance coefficients and zero bit to discard the remaining coefficients can only achieve a compression ratio of 2 at the likely penalty of severe degradation in picture quality.

A Handamard transform system is simulated on a PDP-15 computer at Carleton University whose block diagram is shown in Figure 2. Four different picture samples - two typewritten texts, a line drawing and a circuit diagram - are used, each of size 8"x1" corresponding to 512x128 pixels. Here, each subpicture of 4 x 4 pixels was rearranged as one dimentional data, i.e. as a 16 - component vector f with each component taking on values 1 and -1 corresponding to black pixel and white pixel respectively. With subjective judgement of image quality degradation after each processing, we found that: i) at least seven Hadamard coefficients must be retained exactly to reproduce pictures with barely acceptable degradation; ii) among several coefficient combinations we have tried, reconstructions with the largest variance coefficients yielded the least distorted pictures; and iii) further subjective distortion prevails when a quantization scheme was applied. The circuit diagram picture sample and its reconstructions from 1 and 5 nonquantized and from 6 quantized transform coefficients are shown in Figures 3a, b, c & d respectively. Likewise, Figures 4a & b show a text sample and its reconstruction from 5 nonquantized transform coefficients. The quantization scheme for the 6 transform coefficients of Figure 3d is given in Table 1.

Although we have sofar demonstrated the inefficiency of the transform coding method in the compression of binary pictorial data, it is still hopeful that there is a transform coding scheme which provides some modest data compression (e.g. CR=2) such that it can be used in a hybrid coding system to result in an attractive overall data compression ratio. We are thus motivated to investigate the hybrid technique which combines an original domain coding system and a transform one.

IV. THE HYBRID TREE-TRANSFORM CODING SCHEME

In this approach, the picture is first partitioned into several small equal-size subpictures. A tree sheme is then applied, which assigns short codewords to a few frequently occured subpictures and the shortest codeword to any of the remaining subpictures, which is further coded using a transform coding scheme. For example, a subpicture can be classified as either a "DC" subpicture if it contains all white pixels, or a "non-DC" subpicture if it contains one or more black pixels, and one bit can be used to distinguish between a DC subpicture and a non-DC subpicture. Indeed, a DC subpicture, which occurs very frequently in a picture, can be simply coded by a O and a non-DC subpicture can be coded by a 1 followed by an appropriate codeword of a transform coding scheme.

let N denotes the number of pixels in each subpicture;

n, the number of subpictures in the picture;

M, the total number of bits required to code the picture;

K, the length of subpicture codeword of the transform scheme used;

n_b, the number of non-DC subpictures in the picture;

and r_b, the ratio $\frac{n_b}{n}$ of non-DC subpictures over all subpictures.

Then the compression ratio (CR) of the above hybrid scheme can be expressed as:

$$CR = \frac{Nn}{M} = \frac{Nn}{Kn_b + n} = \frac{N}{Kr_b + 1}$$

Clearly, the efficiency of the hybrid scheme depends on the subpicture size, the efficiency of the transform coding scheme and, of course, the picture statistics. A transform scheme of modest efficiency (e.g. CR=2) can thus be used in this hybrid scheme to achieve an attractive overall compression ratio (e.g. CR=6.15 if N=16, K=8 and r_b=0.2)

It is interesting to note that this hybrid method can be generalized by combining a tree scheme with any block coding scheme such as a transform scheme, a sequence indexing scheme [7], or a scheme using an error correcting code.

V. SIMULATION AND RESULTS

To investigate the feasibility of the transform method in the coding of graphic images whether applied singly or in combination with a tree scheme we have performed a simulation experiment on the Image-100 image processing system at the Canada Center for Remote Sensing (CCRS). Figure 5 shows the block diagram of this simualtion system. Four picture samples (a map of north pole, a text, a circuit diagram, and a handwritten picture) were scanned by the PDS microdensitometer at the resolution of 5, 6, 3 and 2.5 lines per mm respectively. Corresponding pictorial data were then stored on magnetic tape, and later read into the solid-state Intel memory for display on a CRT terminal and for processing by the Image-100 analyser and the PDP 11/70 computer. In addition, we also processed a text sample, shown in Figure 8a, which was generated by the character generator of the Image-100 system.

In the simulation experiments, a Hadamard transform coding scheme with subpicture size 1x16 was first applied singly to the five picture samples, and then in combination with a simple tree scheme which filtered out all DC subpictures. The plots of the calculated variances of the 16 Hadamard coefficients show that energy distribution over the Hadamard transform coefficients varies from picture to picture and that picture energy does not concentrate mostly in the lower order coefficients as was the case for multi-level imagery. As expected, the first (DC) coefficient contains most energy for the case of simple transform coding (Figure 6). But as DC subpictures are removed from the sample, the first Hadamard coefficient collects little energy and the variances of the remaining coefficients become very close to one another, as shown in Figure 7. This rather uniform distribution of picture energy over the transform coefficients clearly indicates the disadvantage of going to the transform domain to code graphic images. Indeed, severe degradation in picture quality was noted from pictures reconstructed from various sets of Hardamard transform coefficients and in general, a large number of coefficients, as many as 13, must be retained exactly to reproduce picture of barely acceptable quality. Figures 8 shows 3 original picture samples - a text, a map, and a circuit diagram - and their reconstructed versions from 13 "exact" transform coefficients. Likewise, Figure 9 shows the degradation of the text picture reconstructed from 13 "exact" coefficients in the hybrid tree - Hadamard transform scheme (in which DC subpictures are filtered out).

We are thus led to believe that the Hadamard transform coding system with subpicture size 1x16, whether applied singly or in combination with a tree scheme, is not feasible for the coding of graphic images.

VI. CONCLUSIONS

I have reviewed various linear transformations for multilevel image coding, and demonstrated, by computer simulation, the infeasibility of the Hadamard transform coding system with subpicture sizes 1x16 and 4x4 in the compression of binary pictorial data, thereby suggesting the impracticability of the linear transform technique in the coding of graphic images in general. Although the performance of the class of linear transform coding schemes for graphic images may be improved by using a different subpicture size, e.g. 2x8, 2x16, 4x4, or 4x8 and a different linear transformation such as the Fourier or the lower triangular, I believe that such improvements, if any at all, would not be so great as to turn a deficient coding scheme into an efficient one.

REFERENCES

[1] Wintz, P.A., "Transform picture coding," Proc. IEEE, vol. 60, p.809 (1972).

[2] Andrews, H.C, Prat, W., and Kaspari, K. "Computer techniques in image processing," Academic Press, New-York (1970).

[3] Pratt, W., Kane, J., and Andrews, H., " Hadamard transform image coding," Proc. IEEE, vol. 57, p.58 (1969).

[4] Landau, H.J., and Slepian, D., "Some computer experiments in picture processing for bandwidth reduction," Bell Syst. Tech. J., vol. 50 (1971).

[5] Habibi, A., "Hybrid Coding of Pictorial Data," IEEE Trans. Communication, vol. 22, p. 614 (1974).

[6] Martin R.S. and Wilkinson J., "Symmetric Decomposition of a Positive Definite Matrix," Numerische Mathematik, p. 355 (1965).

[7] Vuong, S.T., "The Transform Technique and a Tree-Indexing Coding Scheme for The Compression of Binary Pictorial Data," Canadian Communications and Power Conference, Montreal, pp.465 - 568 (1980).

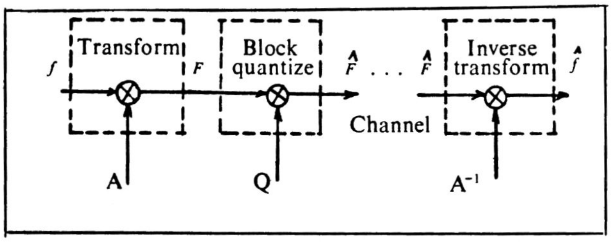


Figure 1. Block diagram of a transform system

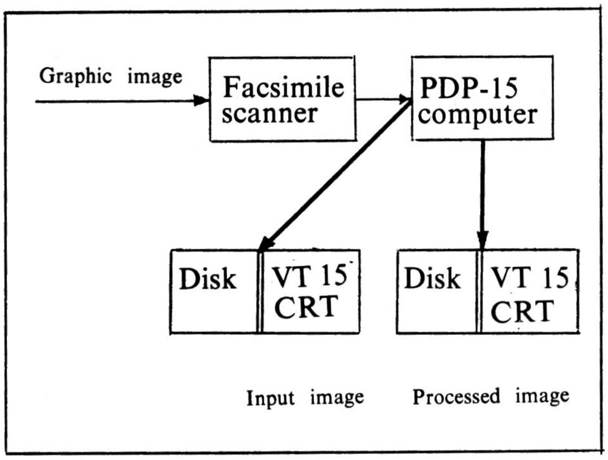
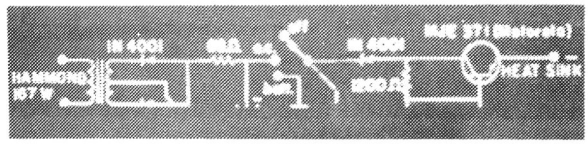


Figure 2. Block diagram of the simulation system at Carleton university

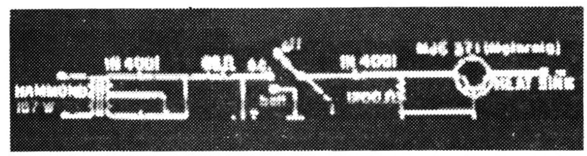
3 bits	0		2		6		10		16	← cut points
		1		5		9		14		← representative points
2 bits	0				6				16	
			3				12			
1 bit	0								16	
					8					

Table 1. Quantization scheme for Figure 3d (similarly for negative value)

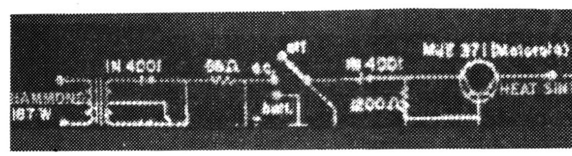


a) Original sample

b) Retaining one "exact" coefficient (DC)

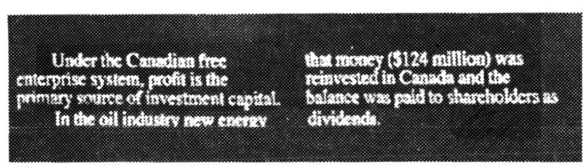


c) Retaining 5 "exact" largest variance coefficients

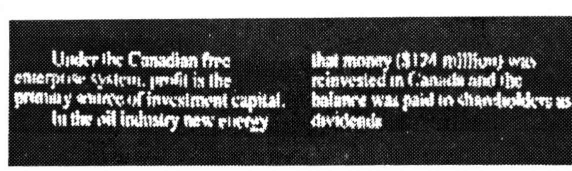


d) Retaining 6 quantized largest variance coefficients

Figure 3. A circuit diagram picture sample and its reconstructions by retaining some transform coefficients



a) Original sample



b) Reconstructed from 5 "exact" Hadamard coefficients

Figure 4. A text picture sample and its reconstructed version

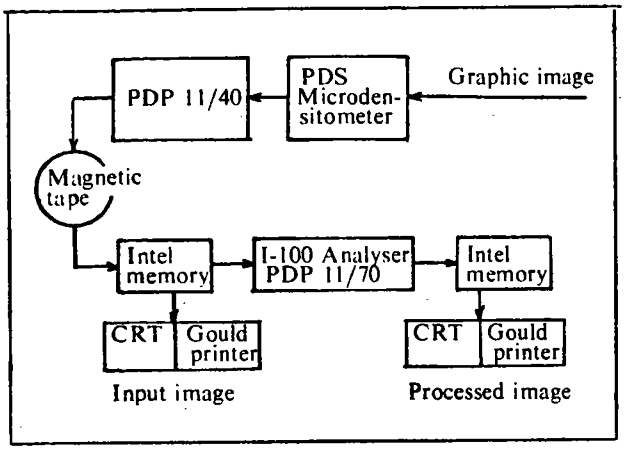


Figure 5. Block diagram of the simulation system at CCRS

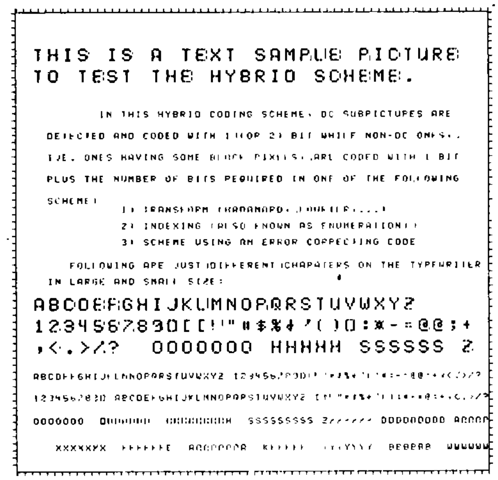


Figure 9. Text picture reconstructed from 13 "exact" coefficients in the tree - Hadamard transform scheme (DC subpictures filtered out)

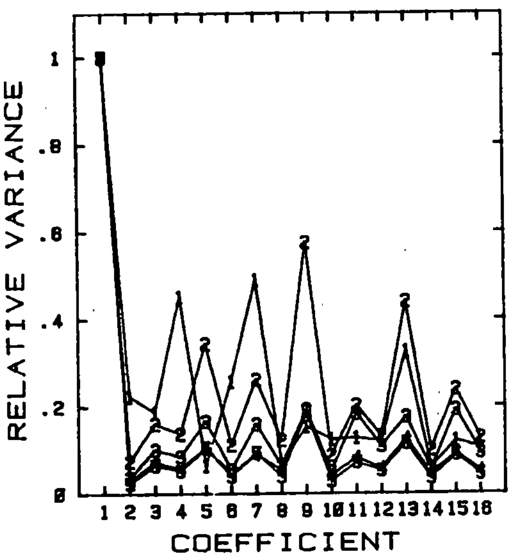


Figure 6. Relative variance of Hadamard transform coefficients for 5 picture samples

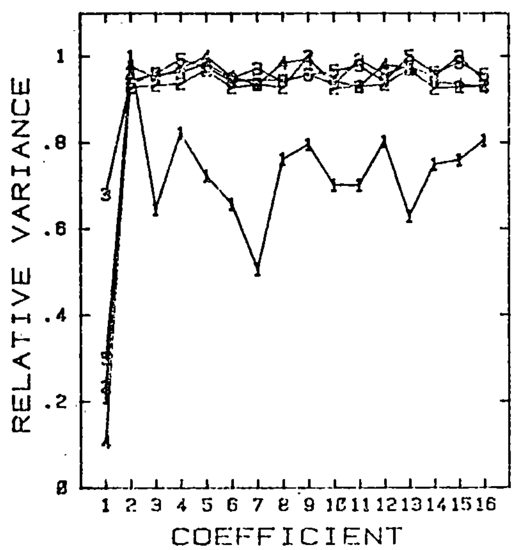
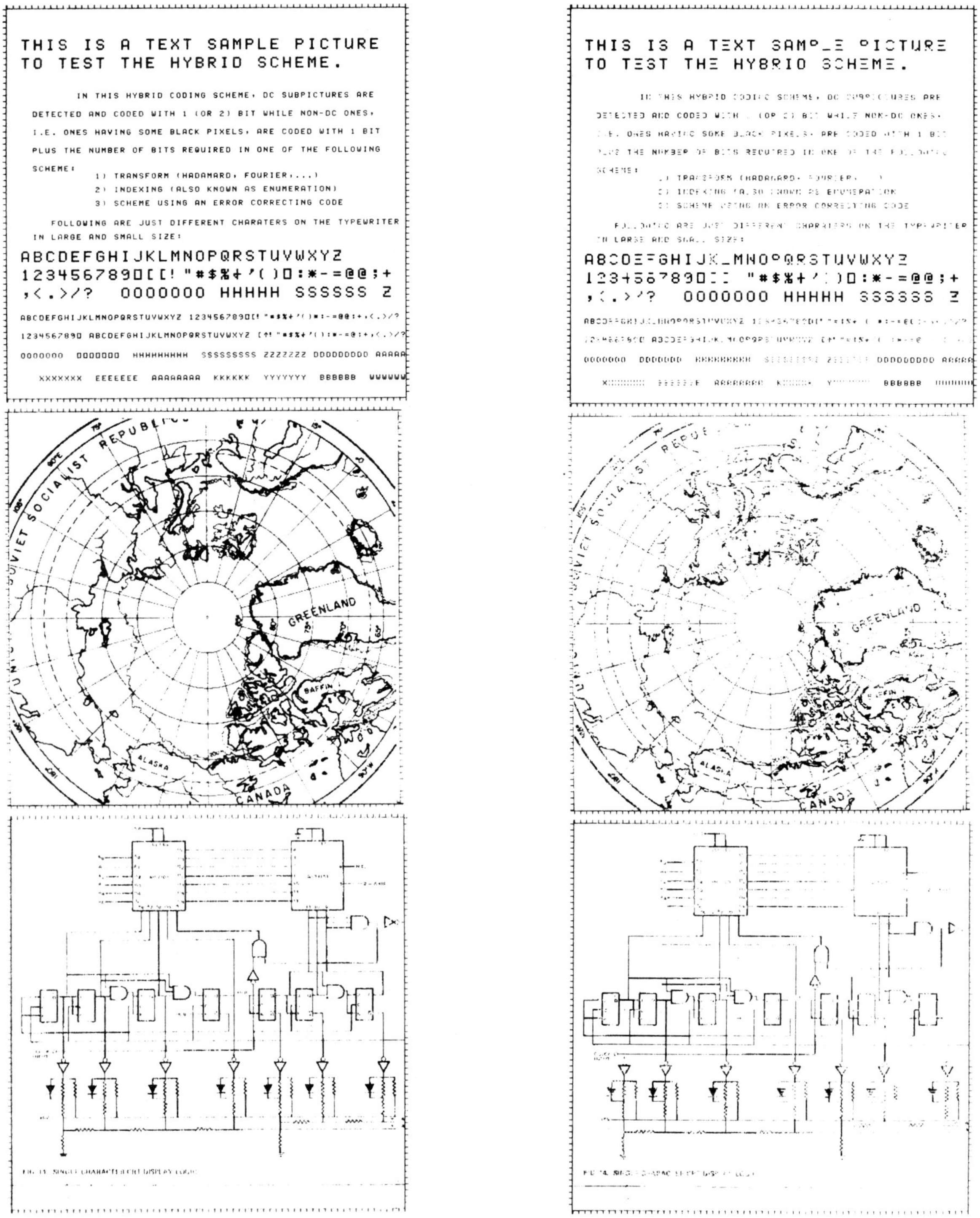


Figure 7. Relative variance of Hadamard transform coefficients for 5 picture samples (DC subpictures filtered out)



a) Original

b) Reconstructed from 13 Hadamard coefficients

Figure 8. Three picture samples and their reconstructed versions from 13 "exact" Hadamard transform coefficients