An Energy Spreading Transform Approach to Efficient Image Retrieval

Waranan Saengow and Bundit Thipakorn
Computer Engineering Department
King Mongkut's University of Technology Thonburi
Thung Kraru District, Bangkok 10140, THAILAND
E-mail: s1410038@cc.kmutt.ac.th
bundit@cpe.eng.kmutt.ac.th

Douglas Cochran
Department of Electrical Engineering
Arizona State University
Tempe, AZ 85287-7206, USA
E-mail: cochran@asu.edu

Abstract

When searching a large database of images for ones that are visually similar to a particular reference images, it is desirable to have efficient means for making coarse comparisons between images. This paper describes an approach for efficient image retrieval using a recently discovered energy-spreading Hadamard transform arising from an orthogonal basis of unimodular sequences called the Prometheus Orthonormal Set (PONS). The energy-spreading properties of this PONS transform suggest that global comparison of key features of images can be accomplished using any subset of the PONS transform coefficients.

1. Introduction

When searching a large database of images for ones that are visually similar to a particular reference image, it is desirable to have a highly efficient means for making coarse comparisons that also scales naturally to support finer comparisons of images which satisfy initial coarse screening criterion.

Crucial information for visual recognition of many images includes both spatially-concentrated structures (e.g., edges) and a frequency-concentrated (e.g., lowpass) background. Recent work by Byrnes et al. [1, 2] has yielded an orthogonal basis for complex n-dimensional space consisting of real unimodular (i.e., ± 1) sequences of length $n=2^m$ having frequency spectra that are optimally flat in several mathematically precise senses. Associated with this basis, called the Prometheus Orthonormal Set (PONS), is a Hadamard transform [3, 4, 5] known as the PONS transform. Because the PONS sequences are unimodular, their energy is spread uniformly with respect to time. Together with the spectral flatness properties of the sequences, this implies the PONS transform provides spreading of both time-concentrated and frequency-concentrated signal energy. The PONS transform hence spreads energy from both critical types of structures used in image recognition among all the transform coefficients, suggesting that information crucial to image recognition will be contained in each transform coefficient.

The approach described in this paper is similar to other recently described work (e.g., [6, 7]) in that it seeks to accomplish image comparison in a transform domain. Its departure from existing approaches is in the use of an energy-spreading transform for the purpose of ensuring that each transform coefficient contains essential information about both spatially-concentrated and frequency-concentrated characteristics of the image that play essential roles in recognition.

2. Hadamard Transforms and The PONS construction

A Hadamard matrix H_n of order n is an $n \times n$ matrix containing only the values $+1$ and -1 and with the property that inner product of any two distinct rows is zero. Thus

$$H_n H_n^T = H_n^T H_n = nI_n \qquad (1)$$

where I_n denotes the $n \times n$ identity matrix. The rows of a Hadamard matrix define a collection of n orthogonal discrete-time signals, each of length n. Moreover, each of these signals has energy n, so the rows of a Hadamard matrix are an basis for C^n that is orthonormal up to the factor $n^{1/2}$ and the linear transformation $x \rightarrow H_n x$ is unitary up to this same constant factor.

The z-transform of each of the rows of a Hadamard matrix is a polynomial of degree n-1 in z^{-1} whose values on the unit circle constitute the discrete-time Fourier transform (DTFT) frequency spectrum of the signal.

A construction due to H.S. Shapiro (see [2]) defines a pair of polynomials $P_n(z)$ and $Q_n(z)$ of degree n-1 with n=2^m having ± 1 coefficients and which are optimally flat on the unit circle; i.e., $\max|P_n(e^{i\omega})|$ and $\max|Q_n(e^{i\omega})|$ both achieve a sharp lower bound that applies to all unimodular polynomials of degree n-1. The construction is inductive with

$$P_0(z) = Q_0(z) = 1 \tag{2}$$

and, for m ≥ 0,

$$\begin{aligned} P_{m+1}(z) &= P_m(z) + z^{-2^m} Q_m(z) \\ Q_{m+1}(z) &= P_m(z) - z^{-2^m} Q_m(z) \end{aligned}$$

PONS is a full basis of ± 1 sequences whose frequency spectra satisfy the same flatness condition as the sequences arising from Shapiro's construction. Byrnes' construction of PONS [2] proceeds using the inductive method based on the concatenation rule depicted by

$$\begin{bmatrix} A \\ B \end{bmatrix} \rightarrow \begin{bmatrix} A & B \\ A & -B \\ B & A \\ B & -A \end{bmatrix} \tag{5}$$

To illustrate the construction, the 2×2 PONS matrix is given by

$$P_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

A 4x4 PONS matrix is constructed using the first row of P_2 as A, and the second row of P_2 as B in the concatenation rule. Similarly,

$$P_4 = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}$$

P_8 is constructed using the first two rows of P_4 and the concatenation rule. Thus

$$P_8 = \begin{bmatrix} 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 \end{bmatrix}$$

The first two rows of P_8 and the concatenation rule (5) yield the first four rows of P_{16}, and so on.

Before proceeding to describe the image retrieval application of the PONS transform, a comment about the role of Hadamard matrices and their associated linear transforms in signal processing are appropriate. First, the use of Hadamard techniques in applications has a well established history [3,4,5] and, in particular, the Walsh transform is a special case that plays important roles in signal processing and communications. The Walsh sequences, however, do not enjoy the same spectral flatness properties as the PONS sequences; the Walsh transform will not generally spread temporally localized energy as evenly among transform coefficients as the PONS transform will. Several alternative PONS constructions are known to Byrnes and his collaborators. But, to the authors' knowledge, none of these are yet published.

3. Approach

For the purposes of this work, issues of scale, rotation, and shift alignment are neglected in favor of developing the fundamental approach for efficient image matching. Under the assumption that two monochrome images I and J of the same size have both been placed in some standard alignment in scale, rotation, and shift, they can be compared by a correlation coefficient

$$\gamma(I,J) = \frac{\langle I,J \rangle}{\|I\|\ \|J\|} \tag{6}$$

In this expression, $\langle \cdot,\cdot \rangle$ donates inner product and $\|\cdot\|$ denotes norm, both in the usual Euclidean sense. The Schwarz inequality implies $-1 \le \gamma \le 1$ with $\gamma = 1$ if and only if $I = \alpha J$ for some positive constant α. Hence computation of γ provides a means for comparing the similarity of images in which (a) $\gamma = 1$ indicates a perfect match up to an intensity scaling and (b) values of γ near unity indicate better matches than smaller values.

Since the rows of a $n\times n$ Hadamard matrix (and the PONS transform matrix P in particular) are orthogonal and all have norm $n^{1/2}$, Parseval's relation implies

$$\gamma(I,J)=\frac{\gamma(PI,PJ)}{\sqrt{n}} \qquad (7)$$

For example, the image correlation can be carried out in the transform domain.

In the proposed approach, images of standard size are stored in the database in the form of PONS transform coefficients of a block decomposition. For instance, a 256x256-pixel image might be decomposed into 16x16 blocks, each of size 16x16 pixels, and the 256-point PONS transform of each block stored in the database. Note that multiplication by a real Hadamard matrix (or its inverse) requires only adds and subtracts (no multiplies) so that formatting images for storage in this form can be accomplished with high computational efficiency.

As already noted, the comparison statistic γ can be computed in the transform domain. But this provides no computational advantage over computing it from the original image data. The viability of this idea is illustrated in the following section of this paper.

4. Reduced Image Representation By PONS Coefficients

To illustrate how single PONS transform coefficients tend to capture both spatially concentrated and frequency concentrated information important to image recognition, this section presents an example of representing an image using a single *randomly chosen* PONS coefficient from each 16x16-pixel block of an image. The 256x256-pixel monochrome image shown in figure 1 was partitioned into 16x16 blocks and both the discrete cosine transform (DCT) and PONS transform of each block were computed. A reconstruction of the image using only the DC coefficient from the DCT in each block is shown together with two reconstructions each using only a single randomly chosen PONS coefficient from each block. Note that the recognizablity of visual features is approximately equal in the PONS and DC reconstructions, though the DC coefficients from each block are the "optimal" choice for reconstruction from DCT coefficients in the sense that they contain the most signal energy. In fact, though not readily visible in this example, the mathematics suggests that the presence of edges and other high-frequency features should be better captured in the PONS coefficients than the DC values.

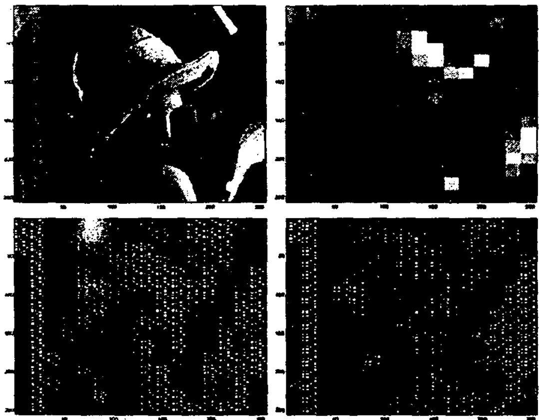

Figure 1: (Upper left) Original image. (Upper right) Image reconstructed using only the DC coefficient from the DCT in each block. (Lower left) Image reconstructed using only a single randomly chosen PONS coefficient from each block. (Lower right) Image reconstructed using a different single randomly-chosen PONS coefficient from each block.

5. Experimental Results

To illustrate the proposed approach to image retrieval, experiments were performed using a single uncategorized database containing 450 images each of size 256x256 pixels. The images were originally 256 gray-scale level 400 Tiff images, but they were stored in this database in the form of PONS transform coefficients from a 16x16 block decomposition. Computations were done in MATLAB.

A query image was transformed into PONS coefficients in the same manner and the database was searched by computing the correlation coefficient defined in equation (6) between the query image and each image in the database. However, instead of using all transform coefficients from each image in computing the correlation, only one coefficient from each block was used (i.e., the correlations were of size 256 rather than 65,536).

Examples of query images together with their best matches from the database are shown in figures 2 and 3. Figure 2 shows the query image (correlation value 1.0000) and eight other images with correlation coefficients (γ) of 0.9756, 0.9733, 0.9731, 0.9721, 0.9707, 0.9683, 0.9621 and 0.9587, respectively. Note the visual similarity of these images with the query image compared to, say, the images in figure 3 – which were also in the database.

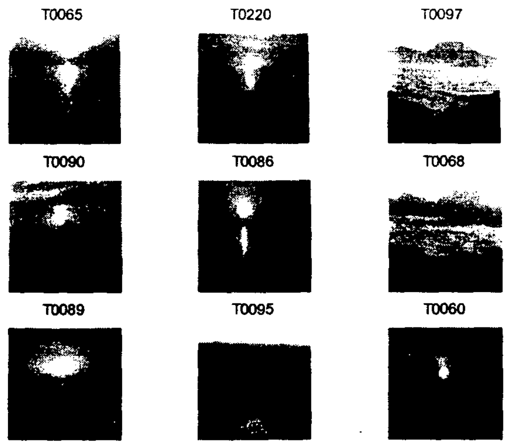

Figure 2: Query image (upper left) and images retrieved from a reduced-order correlation of PONS coefficients.

Similar results are shown in figure 3, where a query image and five retrieved images are with correlation values of 0.9917, 0.9915, 0.9898, 0.9883 and 0.9880, respectively.

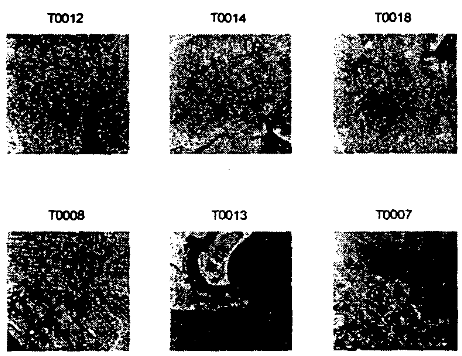

Figure 3: Query image and five retrieved images. Note the qualitative visual similarity of the images to each other as compared, for example, to those in figure 2 (which were also in the query database).

6. Conclusions

We have proposed the use of a vector containing a single PONS coefficient from each block of a block-transform representation as the feature vectors for query retrieval of images similar to a query image from a database. This small feature vector allows efficient coarse comparison of images by correlation in a computationally efficient manner. Moreover, the properties of the PONS transform imply that both low-frequency and high-frequency information that is important in visual recognition of images will be represented in the feature vectors, and hence relected in the correlation comparison of images.

While there are obvious drawbacks to storing image data in the form of transform coefficients, we note that both the PONS transform and its inverse can be computed without multiplies. Thus a front-end to the such a database could be implemented with high computational efficiency.

In future work, we plan to try this method for retrieval of medical images.

References

[1] J.S. Byrnes, B. Saffari, and H.S. Shapiro, "Energy Spreading and Data Compression using the Prometheus Orthonormal Set," *Proceedings of the IEEE Digital Signal Processing Workshop*, Leon Norway, pp. 9-12, September 1996.

[2] J.S. Byrnes, "Quadrature Mirror Filters, Low Crest Factor Arrays, Functions Achieving Optimal Uncertainty Principle Bounds, and Complete Orthonormal Sequences – A Unified Approach," *Applied and Computational Harmonic Analysis*, vol. 1, no. 3, June 1994.

[3] R.K. Rao Yarlagadda and J.E. Hershey, *Hadamard Matrix Analysis and Synthesis with Applications to Communications and Signal/Image Processing*. Kluwer Academic Publishers, 1997.

[4] S.S. Agaian, *Hadamard Matrices and their Applications*. Lecture Notes in Mathematics No. 1168, Springer Verlag, 1980.

[5] M. Harwit and N.J.A. Sloane, *Hadamard Transform Optics*. Academic Press, 1979.

[6] J.A. Lay and L. Guan, "Image retrieval based on energy histograms of the low frequency DCT coefficients," *Proceeding of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 3009-3012, 1999.

[7] Y.L. Huang and R.E. Chang, "Texture Features for DCT-coded image retrieval and classification," *Proceeding of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 3013-3016, 1999.