IMAGE CODING BASED ON VISUAL VECTOR QUANTIZATION

G Qiu*, M R Varley* and T J Terrell*

* University of Derby, UK.
+ University of Central Lancashire, UK.

ABSTRACT

In this paper, the authors present a novel image coding technique which explicitly aims to retain perceptually important features of the image. The system exploits the psychovisual properties of the human visual system, in that it identifies perceptually important vector patterns (referred to as visual vectors) within small blocks of the image, and codes these using a scheme similar to vector quantization (VQ). Thus the system is referred to as visual vector quantization (VVQ).

One of the main advantages of the new VVQ scheme over traditional VQ is that it is computationally much simpler, and yet its performance is comparable to that of traditional VQ. Experimental results are presented which demonstrate that the new technique can achieve visually satisfactory image reconstruction at a compression ratio as high as 17:1. Comparison of peak signal to noise ratio (PSNR) values shows that the performance of the new technique is comparable to that of the traditional VQ-based image coder described by Ramamurthi and Gersho [2] at a similar bit rate, whilst the computational complexity is significantly reduced.

INTRODUCTION

Traditional vector quantization (VQ) [1] for image coding operates in ‘pixel space’. The image to be coded is divided into small blocks, and the pixel values within a block form a multi-dimensional vector. It is well known that the encoding process for traditional VQ-based image compression schemes is computationally intensive.

In this paper, a new image coding system is presented, in which perceptually important features of the block (the horizontal and vertical derivatives, i.e. quantitative measures of horizontal and vertical edge information within the block) are extracted and used to form a ‘visual vector’. An image block is coded by classifying each of these features individually. Thus coding consists of two sets of one-dimensional tests, reducing the computational load of coding significantly. The system codebook is created by training a neural network with the visual vectors as input, and the residual training blocks as the desired output. Residual block reconstruction is achieved by substituting the appropriate image block from the codebook, and the image block is then reconstructed by adding the quantized block mean value to each reconstructed residual pixel.

Experimental results using a block size of 4 pixels x 4 pixels are presented, which demonstrate that this new technique can achieve perceptually satisfactory reconstructed images following coding, with a compression ratio as high as 17:1. Comparison of peak signal to noise ratio (PSNR) values shows that the performance of the new technique is comparable to that of the traditional VQ-based image coder described by Ramamurthi and Gersho [2] at a similar bit rate, whilst the computational complexity is significantly reduced.

SYSTEM CODEBOOK DESIGN

The system codebook is created in two stages, each using data from selected training image(s). Firstly, the training data is segmented into blocks of size m x n pixels, and a ‘visual vector’ for each training data block is found and used to classify the training blocks into nine separate classes, corresponding to different edge information content. Secondly, a neural network is trained to reproduce the residual block pixel values for each of the nine classes, and the outputs of the neural network (m x n reconstructed residual pixel values) are taken as the codebook entry for that class of block.

Block Classification by Visual Vector

The horizontal and vertical derivatives of an image block, $D_h$ and $D_v$, are found according to equations (1) and (2) (equations are for $m = n = 4$).

\[
D_h = \frac{1}{8} \sum_{j=0}^{3} \sum_{i=0}^{3} I(i,j) G_h(i,j)
\]  

\[
D_v = \frac{1}{8} \sum_{j'=0}^{3} \sum_{i'=0}^{3} I(i,j) G_v(i,j)
\]
where \( R(i,j) \) is the pixel value in the block, and \( G_h(i,j) \) and \( G_v(i,j) \) are given by:

\[
G_h(i,j) = \begin{bmatrix}
1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 \\
\end{bmatrix}
\]

\( G_v(i,j) = \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
-1 & -1 & -1 & -1 \\
-1 & -1 & -1 & -1 \\
\end{bmatrix}
\]

This results in a pair of values for each block, i.e. a data set in the two-dimensional 'visual vector' space. A competitive learning algorithm described in [3] is used to cluster the \( D_h \) values for the training blocks into several classes. For the results presented below, three classes were used corresponding to the directional derivative being negative large (NL), small (S) or positive large (PL). This classification process was repeated for the vertical derivative \( D_v \). Since, for the results presented below, three classes were defined in each direction, an image block may fall into one of nine different classes. The three cluster centre co-ordinates for each direction are stored, and used during the coding stage to classify an image block into one of the nine classes.

**Codebook Design by Neural Network**

The system codebook entry for each class of block is found by training a backpropagation neural network, as described by Rumelhart et al. [4], to reproduce the residual block \( m \times n \) pixel values) at its output, when the \( D_h \) and \( D_v \) co-ordinates for the appropriate class are presented at its input. After the network has been trained to reproduce all nine classes of block, the network outputs for each of the nine classes are taken as the codebook entries. Thus the codebook needed to reconstruct the residual blocks consists of nine entries, each of which comprises \( m \times n \) pixel values.

**IMAGE CODING BASED ON VVQ**

The VVQ image coding system operates as indicated in Fig. 1.

The visual feature extractor calculates the directional derivatives of the block, \( D_h \) and \( D_v \), according to equations (1) and (2). The block mean value is also calculated, and the quantized block mean constitutes part of the compressed data for the block. The directional derivatives, \( D_h \) and \( D_v \), are used by the visual feature classifier to classify the block into one of the nine classes. This classification process is simple since only six comparisons are required (three for \( D_h \) and three for \( D_v \)). A class index denoting the class of the block, and the quantized mean value of the block, constitute the compressed data for the block.

Reconstruction of the block is achieved by substituting the appropriate image block from the codebook, which resides at the reconstruction stage of the system, and adding the quantized block mean to each pixel value. Note that the system codebook is not used during the image coding stage - instead an image block is classified according to its calculated \( D_h \) and \( D_v \) values.

**EXPERIMENTAL RESULTS**

To evaluate the VVQ method, the authors used two 512 \( \times \) 512 monochrome images with 8 bits per pixel. One image (F16) was used to train the system as described above, and the other (Lena) was used to test the system. A block size of \( 4 \times 4 \) pixels was chosen, and thus the training data set consisted of 16384 blocks.

For the training data used, the \( D_h \) values were clustered into three groups, the centres of which were found to be located at \(-27\) (NL), \(-0.3\) (S) and \(31\) (PL). Similarly the \( D_v \) cluster centres were found to be at \(-20\) (NL), \(0.1\) (S) and \(25\) (PL). The nine block classes are obtained by combining these \( D_h \) and \( D_v \) values as depicted in Fig. 2.

The system codebook obtained by training the neural network for all the nine classes is shown in Fig. 3. For display purposes, a constant value of 127 has been added to each pixel. The edge content for each of the nine different classes of block is apparent in the codebook.

In the tests, it was found that a large majority of the blocks fell into class 5 (no significant edge content), and hence a variable bit rate coding strategy was used. One bit was used to indicate whether a block belongs to class 5 or one of the other eight classes. Thus, if a block belongs to class 5, one bit is used for the class index, and if it belongs to one of the other eight classes, four bits are used. The block mean was coded using 6 bits. A three layer neural network with 2 input neurons (corresponding to the \( D_h \) and \( D_v \) values for a block), 15 hidden and 16 output neurons (corresponding to the reconstructed residual block pixel values) was used to create the system codebook.

Fig. 4(a) shows the original Lena image, and Fig. 4(b) shows the reconstructed Lena image at a bit rate of 0.47
bits per pixel. The PSNR value for this image, calculated using equation (5), was found to be 29.59 dB.

$$\text{PSNR} = 10 \log_{10} \left( \frac{255^2}{MSE} \right) \text{ dB} \quad (5)$$

where

$$MSE = \frac{1}{N} \sum_{i=0}^{N-1} [x(i) - x_r(i)]^2 \quad (6)$$

and $N$ is the number of pixels in the image, $x(i)$ represents the original pixel values, and $x_r(i)$ represents the reconstructed pixel values.

It can be seen from Fig. 4 that the visual quality is satisfactory. Comparison of the PSNR performance shows that the technique can achieve comparable results to other image coders. For example, the same image, Lena, coded by the VQ technique of Ramamurthi and Gersho [2] at a bit rate of 0.5 bits per pixel has a PSNR value of 28.55 dB.

For the chosen block size ($4 \times 4$) with a codebook of size nine, the computational complexity of the VVQ technique is reduced by a factor of approximately 24 compared with the traditional full search VQ.

**CONCLUDING REMARKS**

The authors have described a new image coding system which exploits the psychovisual properties of the human visual system to represent an image block by a 'visual vector'. It has been demonstrated that this technique can achieve comparable results to traditional VQ-based techniques, with a significant saving in computational load. Results presented demonstrate satisfactory image reconstruction at a compression ratio of 17:1, with a significant reduction in computational complexity.

**ACKNOWLEDGEMENT**

The authors are grateful to Mr. Andrew Stewart for his help in obtaining and reproducing the images in this paper.

**REFERENCES**


<table>
<thead>
<tr>
<th></th>
<th>NL</th>
<th>S</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL</td>
<td>(-27, 0)</td>
<td>(-27, 0.1)</td>
<td>(-27, 0.25)</td>
</tr>
<tr>
<td>Class 1</td>
<td>Class 2</td>
<td>Class 3</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>(-0.3, 20)</td>
<td>(-0.3, 0.1)</td>
<td>(-0.3, 0.25)</td>
</tr>
<tr>
<td>Class 4</td>
<td>Class 5</td>
<td>Class 6</td>
<td></td>
</tr>
<tr>
<td>PL</td>
<td>(31, 20)</td>
<td>(31, 0.1)</td>
<td>(31, 0.25)</td>
</tr>
<tr>
<td>Class 7</td>
<td>Class 8</td>
<td>Class 9</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2 $D_h$ and $D_v$ co-ordinates for the nine block classes

Fig. 3 Visual Vector Codebook
Fig. 4(a) Original Lena Image, 8 bits per pixel

Fig. 4(b) Reconstructed Lena image
Bit rate of coded image = 0.47 bits per pixel, PSNR = 29.59 dB