Using Visual Feature Extraction Neural Network Model To Improve Performance of Quadtree Based Image Coding

Zhongmin He and Sheng Chen

Department of Electrical and Electronic Engineering,
University of Portsmouth, Anglesea Road, Portsmouth PO1 3DJ, U.K.
Telephone: +44-1705-842580, E-mail: zhe@ee.port.ac.uk

Abstract

In this paper, we propose a new technique to improve the performance of quadtree (QT) based image coding through the utilization of a neural network based visual feature extraction model (VFEM). After QT reconstruction is completed, a trained VFEM uses the information contained in the QT reconstructed image to recover the QT reconstruction error. This results in a better quality reconstructed image than the one simply reconstructed from QT representation. Since no extra information other than QT structure itself needs to be transmitted, the VFEM improvement does not increase the coding bit rate. Therefore, a better rate-distortion performance is achieved.

1 Introduction

Neural networks have widely been used for image coding applications and considerable research progresses have been made in this area. In most cases, neural networks are used as part of coding systems to replace some traditional techniques [1]. For example, multilayer perceptrons are used as nonlinear predictors in predictive coding [2] [3], Hebbian networks are used to perform the Karhunen-Loève transform (KLT) in transform coding [4] [5] and Kohonen self-organizing feature maps (KSOFM) are used to design codebooks in vector quantization [6] [7].

In this paper, we introduce a new way of employing neural networks to improve image coding performance. In the proposed method, a neural network model based on some characteristics of visual cortex, referred to as visual feature extraction model (VFEM), is added to a standard image coding system. The VFEM exploits the underlying physics between a reconstructed image and its reconstruction error image (the error between the reconstructed image and the original image), and is trained to generate the reconstruction error image from the reconstructed image. The output of the VFEM is added to the reconstructed image to produce a better quality output image. Since the input to the VFEM is the reconstructed image, no extra information needs to be transmitted and, therefore, a better quality image is obtained without increasing coding bit rate. This means that a better rate-distortion (R-D) performance is achieved.

Hierarchical data structures are important representation techniques in image processing. Quadtree (QT) data structure belongs to a class of hierarchical data structures which are based on the principle of recursive decomposition of image data space [8]. Since QT was introduced into image representation, many research efforts have been made to enhance the QT algorithm itself [9] [10]. In this paper, we use the VFEM to improve the performance of QT coding without altering the QT coding system at all. A VFEM is connected to a QT coding system and is trained to generate the error image of QT reconstruction. The output of the VFEM is added to the QT reconstructed image to recover the QT reconstruction error. The proposed improving method produces a better quality reconstructed image without requiring any extra coding bit rate as the input to the VFEM is supplied by QT reconstruction.

The rest of the paper is organized as follows. Section 2 describes the structure of VFEM and the QT decomposition algorithm is summarised in section 3. Section 4 discusses the application of VFEM to improve QT coding performance. Simulation results are presented in section 5 and some concluding remarks are given in section 6.

2 Visual Feature Extraction Neural Network Model

Artificial neural netwroks have been studied for many years in the hope of achieving "human-like" performance in the field of image processing. According to the understanding of visual cortex [11], there are certain types of feature detectors in the visual cortex which are labelled as *simple cells* and *complex cells*. Simple cells do not respond to diffuse illumination but respond strongly to stimuli such as bars or edges, and complex cells also respond maximally to stimuli but have larger receptive fields and can generalise their response over a wider area of the visual field.

Edge patterns (visual features) in a visual field can be represented to a certain degree by the first order directional derivatives of different scales (different size blocks). Derivatives of small blocks can represent visual features detected by simple cells, and derivatives of large blocks can represent visual features detected by complex cells. Horizontal and vertical derivatives are normally employed as directional derivatives since they can represent the intensity changes (edges) in each direction and the combination of them can determine orientations of edge patterns.

Based on this knowledge, a neural network based visual cortex model for image reconstruction has been proposed [12], in which directional derivatives are calculated for the residual image which is generated by removing image block means from the image itself. Both derivatives and block means are coded and transmitted. At the receiver end, the residual image is reproduced from the derivatives via the visual cortex model, and is then added to block means to reconstruct the original image.

In this paper, an artificial neural network model with some neurons specially designed to extract the visual features of the input image is presented. The model, referred to as VFEM, is a two-hidden-layer feedforward neural network shown in Fig.1. The first hidden layer of the network is the visual feature extractor which calcultes multi-scale derivatives of the reconstructed image, and the rest of the network uses these visual feature information to generate the error image. In our application, a VFEM is added to a image coding system and is used to generate the reconstruction error image from the reconstructed image supplied by the existing coding system. It should be emphasized that the purpose of the VFENNM is different from that of the visual cortex model presented in [12].

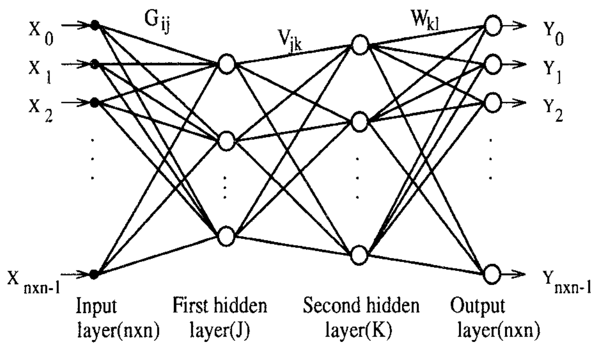

Figure 1: Visual feature extraction neural network model

The input to the VFEM, the reconstrcted image produced by the coding system, is first divided into blocks of size $n \times n$ and then fed into the VFEM. To calculate derivatives of image blocks in different scales, each image block is recursively divided into 4 equal-sized sub-blocks until the sub-block size is reduced to 2×2. For each block or sub-block X of size $n_b \times n_b$, its horizontal derivative d_h and vertical derivative d_v are calculated as:

$$d_h = \sum_{i=0}^{n_b-1} \sum_{j=0}^{n_b-1} x(i,\ j) \cdot g_h(i,j) \qquad (1)$$

$$d_v = \sum_{i=0}^{n_b-1} \sum_{j=0}^{n_b-1} x(i,\ j) \cdot g_v(i,j) \qquad (2)$$

The derivative kernels g_h and g_v can be written collectively in matrix form as

$$G_h = \begin{bmatrix} \mathbf{1} & \mathbf{-1} \\ \mathbf{1} & \mathbf{-1} \end{bmatrix} \qquad G_v = \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{-1} & \mathbf{-1} \end{bmatrix}$$

where the bold numbers $\mathbf{1}$ in G_h and G_v represent $\frac{n_b}{2} \times \frac{n_b}{2}$ matrices of 1s, and the bold numbers $\mathbf{-1}$ represent $\frac{n_b}{2} \times \frac{n_b}{2}$ matrices of -1s.

The first hidden layer of the VFEN acts as a visual feature extractor to calculate block derivatives. Hence the number of neurons in the first hidden layer is equal to the number of total derivatives required for a block, and the weights G_{ij} between the input layer and the first hidden layer are fixed according to equations (1) and (2) to serve the purpose of derivative calculation.

The rest of the network, the second hidden layer and the output layer, are used to reconstruct the error image from the derivative information supplied by the first hidden layer. The number of neurons in the output layer is equal to the number of pixels in an image block ($n \times n$). The appropriate number of neurons in the second hidden layer is decided by experiment. The weights V_{jk} connecting the first hidden layer to the second hidden layer and W_{kl} connecting the second hidden layer to the output layer are learnt via supervised training [13].

3 Quadtree Decomposition

QT decomposition is a simple and efficient technique for image representation at different resolution levels. This representation can be useful for a variety of image processing applications, such as pattern recognition and image compression. QT decomposition divides an image into homogeneous regions of different sizes depending on the activity in each region. Such kind of division makes compression adaptive to various activities of image regions and, therefore, is more efficient than regular one-size-block division.

An image of size $2^N \times 2^N$ can be represented by a QT of $N+1$ levels. The root of the QT is associated with the whole image and the QT representation is built up as follows. If the image or an image block is homogeneous according to a chosen homogeneous criterion, it becomes a *leaf* of the tree; otherwise, it becomes a *node* of the tree and is further divided into four equal-sized sub-blocks (quadrants). Same operations are repeated on sub-blocks until all of them find homogeneity. It is obvious that leaves in different QT levels are associated with image blocks of different sizes and, in particular, leaves in QT bottom level are associated with individual image pixels. The QT decomposition can be done in either top-down or bottom-up manner.

When QT decomposition is completed, the resulting tree is coded for transmission or storage. Both the tree structure information and leaf intensity information need to be coded. At reconstruction stage, the decoded QT structure data and leaf intensity data are used to rebuild the QT representation and then to reconstruct the original image.

Given an image of size $2^N \times 2^N$, let $x_i(k,l)$ be the pixel value at position (k,l) in QT level i and T be the homogeneous threshold. Then the QT decomposition procedure can be summarized as follows [9]:

Step 1: Set $i = 1, T_i = T$.
Step 2: FOR $k, l = 0, \cdots, 2^{N-i} - 1$;
  IF all $x_{i-1}(2k+p,\ 2l+q)$
    for $p, q = 0, 1$ are leaves;
  THEN perform the homogeneity
    test according to equation (3);
    if the test result is TRUE,
     then $x_i(k,l)$ becomes a leaf;
     else $x_i(k,l)$ becomes a node;
  ELSE $x_i(k,l)$ becomes a node;
  NEXT k, l.
Step 3: IF $i > N$, goto Step 4;
  ELSE $i = i+1, T_i = T_{i-1}/2$,
    and goto Step 2.
Step 4: Code tree structure information.
Step 5: Allocate bits for leaves in each level according to equation (5) and quantize the leaves.
Step 6: Stop.

The homogeneity test is defined as

$$Max\ |x_i(k,l) - x_{i-1}(2k+p, 2l+q)| \le T_i \quad for\ \ p, q = 0, 1 \tag{3}$$

where

$$x_i(k,l) = \frac{1}{4}\sum_{p=0}^{1}\sum_{q=0}^{1} x_{i-1}(2k+p, 2l+q) \quad (4)$$

The homogeneous threshold T controls the QT R-D performance trade-off. A smaller threshold results in a more complex tree structure, hence a higher coding bit rate and a better quality reconstructed image, and vice versa.

The number of bits B_i allocated for quantizing leaves in QT level i is

$$B_i = \frac{1}{2}\log\frac{\sigma_i^2 \, L_{qt}}{4^{N-i}\,D} \quad (5)$$

where L_{qt} is the number of total leaves in the QT structure, σ_i^2 is the variance of leaves within QT level i and D is a chosen distortion for quantization.

4 Improving QT Coding Via VFEM

A QT coding system with the VFEM improvement is illustrated in Fig.2. The first two parts of the system in Fig.2 comprise the standard QT coding system which perform QT decomposition and reconstruction. After the QT reconstruction is completed, the QT reconstructed image is passed to the VFEM, and the output of the VFEM, the estimated error image, is added to the QT reconstructed image to produce the final output image.

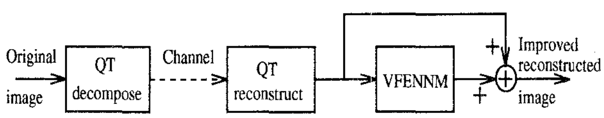

Figure 2: QT Coding system improved by VFEM

Obviously, in order for the VFEM to be able to recover the error image of QT recostruction, it must be designed and trained properly. Training data is comprised of some QT reconstructed images as input to the VFEM and the correponding QT reconstruction error images as desired output for the VFBNN. A QT reconstructed image contains blocks of different sizes, and each block is represented by its mean. Subtracting a QT reconstructed image from its corresponding original image generates the reconstruction error image which contains only edge patterns. Multiscale derivative information calculated from the QT reconstructed image by the first hidden layer of the VFEM can already represent the edge patterns in the error image to a certain degree. The rest of the VFEM is designed to learn a more accurate representation of the underlying correlation between QT reconstructed image and QT reconstruction error image. Because VFEMs are usually large, we adopt the backpropagation algorithm [14] in training to keep a minimum computational requirement.

5 Simulation Results

Simulation was conducted to study the performance of the VFEM and to investigate the effects of image block size, training images and number of neurons in the second hidden layer. Four 512×512 images shown in Fig.3, with 8 bits per pixel, were involved in the simulation either as training images or test images.

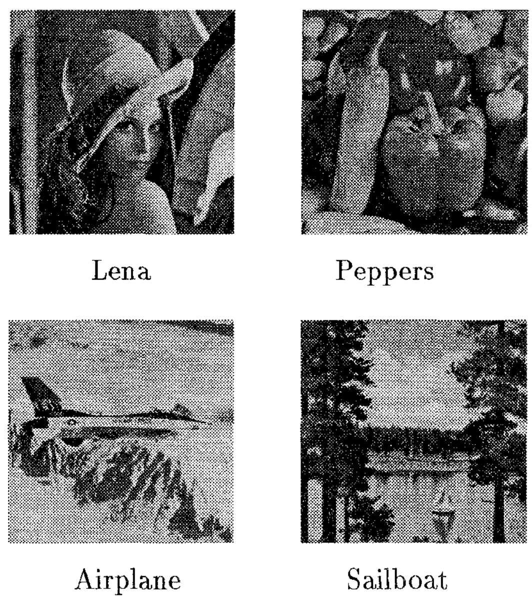

Figure 3: Original images used in simulation

A key parameter of the proposed scheme is the image block size. The design choice for this is a trade-off between performance and complexity. A QT reconstructed image consists of homogeneous blocks (leaves) of different sizes. A QT leaf of certain size can contribute to derivative information only when the image block size is larger than the size of the QT leaf. For example, a QT leaf of size 8×8 can only contribute to the derivative information when image block size is 16×16 or over. In general, increasing image block size can provide more derivative information and thus improve the VFEM's performance, providing that enough training images are available to train the network properly. Otherwise, choosing a large image block size without enough training images can result in poor performance due to insufficient training. It is also obvious that increasing image block size increases the size of the network and consequently the computational load.

The number of neurons in the second hidden layer is decided by experiment. The experiment was started with a small number of neurons in the second hidden layer and gradually more neurons are added until further increase does not lead to any additional improvements of the VFEM performance.

Among the four images given in Fig.3, when one of them was chosen as test or coding image, the other three images were then used as training images. An image block size 16×16 and a 40-neuron second hidden layer was found to be most appropriate. Table 1 gives the performance improvement on QT coding via VFBNN at a coding bit rate of 0.5 bits per pixel (bpp). The peak signal to noise ratio (PSNR) used in Table 1 is defined as

$$PSNR = 10 \cdot \log_{10}(\frac{255^2}{MSE}) \quad \text{dB} \tag{6}$$

where

$$MSE = \frac{1}{4^N} \sum_{i=0}^{2^N-1} \sum_{j=0}^{2^N-1} [x_{ij} - \hat{x}_{ij}]^2 \tag{7}$$

X is the original image of size $2^N \times 2^N$ and $\hat{X}$ is the reconstructed image.

Table 1: Perfomance improvement on QT coding via VFEM (bit rate = 0.5 bpp)

Coding image	PSNR (dB)		
	QT	QT+ VFEM	Improve-ment
Lena	32.38	33.30	0.92
Peppers	33.28	34.15	0.87
Airplane	32.86	33.86	1.00
Sailboat	29.53	30.48	0.95

Fig.4 shows the QT reconstructed image (a) and VFEM improved image (b) of image "Lena".

(a) QT reconstructed (0.5 bpp, 32.38 dB)

(b) VFEMNN improved (0.5 bpp, 33.30 dB)

Figure 4: QT reconstruction and VFEM improvement on image "Lena"

6 Conclusions

A new technique for improving the performance of QT based coding using artificial neural networks has been proposed. The proposed scheme improves the QT reconstructed image quality through the recovering of the QT reconstruction error via a VFEM. It is worthy pointing out that the VFEM is added to a QT coding system without altering the existing system at all. An important advantage of the proposed technique is that the reconstructed image quality is improved while the coding bit rate remains unincreased. Experimental results show that significant improvements were obtained and even better improvement can be achieved if more training images are available. This research opens up a promising future for a new category of techniques to improve image coding performance using the VFEM not only for QT based coding but also for other coding methods.

References

[1] R. D. Dony and S. Haykin, "Neural network approaches to image compression," *Proc. IEEE,* Vol. 83, No. 2, pp.288-303, February 1995.

[2] S. A. Dianat, N. M. Nasrabadi and S. Venkataraman, "A nonlinear predictor for differential pulse-code encoder (DPCM) using artificial neural networks," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing '91,* pp. 2793-2796, May 1991.

[3] N. Mohsenian and S. A. Rizvi, "Predictive vector quantization using a neural network approach," *Optical Engineering,* Vol. 32, No. 7, pp. 1503-1513, July 1993.

[4] T. D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network," *Neural Networks,* Vol. 2, pp.459-473, 1989.

[5] S. Y. Kung, K. I. Diamantaras and J. S. Taur, "Adaptive principal component extraction (APEX) and applications," *IEEE Trans. Signal Processing,* Vol. 42, No. 5, pp. 1202-1217, May 1994.

[6] N. M. Nasrabadi and Y. Feng, "Vector quantization of images based upon the Kohonen self-organizing feature maps," *Proc. IEEE Int. Conf. Neural Networks,* Vol. 1-2, pp. I 101-105, July 1989.

[7] H. Liu and David Y. Y. Yun, "Self-organizing finite state vector quantization for image coding," *Proc. Int. Workshop Applications of Neural Networks to Telecommunications,* pp.176-182, 1993.

[8] H. Samet, "The quadtree and related data structures," *Computing Surveys,* Vol. 16, No. 2, pp. 187-260, June 1984.

[9] E. Shusterman and M. Feder, "Image compression via improved quadtree decomposition algorithm," *IEEE Trans. on Image Processing,* Vol. 3, No. 2, pp. 207-215, March 1994.

[10] L. Berger, J. P. Mariot and C. Launay, "A new formulation for fast image coding using quadtree representation," *Pattern Recognition Letters,* Vol. 13, No. 6, pp. 425-432, June 1992.

[11] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology,* Vol. 160, pp. 106-154, 1962.

[12] G. Qiu, "Modelling the visual cortex using artificial neural networks for visual image reconstruction," *Artificial Neural Networks, Conference Publication,* No. 409, PP.127-132, 26-28 June 1995.

[13] S. Haykin, *Neural Networks: A Comprehensive Foundation.* New York: Macmillan Collage Publishing Company, 1994.

[14] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning internal representations by error propagation," Chapter 8 in *Parallel Distributed Processing.* Vol. 1, Cambridge, MA, MIT Press, 1986.