A Fully Neural Approach to Image Segmentation and Image Coding

Andreas Koenig, Michael Reinke, Manfred Glesner

Institute for Microelectronic Systems
Darmstadt University of Technology
D-6100 Darmstadt, Germany

A fully neural approach to image segmentation, image coding and feature analysis of facial images is presented. An algorithmic quadtree image decomposition method is used to generate training data for a group of multi-layer-perceptrons (MLP), which are trained with a standard error-backpropagation rule. The generated segmented image is equivalent to its quadtree representation and is used as input for following facial feature extraction and variable-block-size vector quantization (VBSVQ) steps. The codebook generation for the VBSVQ image coder and the image coding itself is done by a Kohonen feature map. The extracted facial image features, e.g eyes, nose and mouth, may serve as input data for new coding techniques like semantic coding or techniques based on motion estimation.

1 Introduction

A very important field of digital image transmission is face-to-face image transmission using (digital) telephony lines. Due to the low bitrate capacity of these lines (ISDN 64 kbit/s) sophisticated coding methods are required. The purpose of this paper is to present an image coder based on image segmentation combined with vector quantization in a fully neural environment. Such an approach offers significant advantages both in performance and implementation complexity. An additional application based on the segmentation step is the extraction of important facial features which is done by neural networks, too. This facial feature recognition is done by the investigation of a binary representation of the source image different to the approach of Vincent [4], which uses fine resolution feature locators to investigate the original grayscale picture. The cascadable VLSI-system BACCHUS developed at Darmstadt University of Technology (DUT) seems to be a suitable hardware implementation for the facial feature recognition [1,2,3]. This neuron like binary associative memory serves as hardware implementation in the fields of industrial quality control, handwritten letter recognition and image comression. In prior work an image coding with the procedure of vector quantization has been implemented succesfully [1,2]. The resulted feature representation of the facial image is the starting point for the recognition of the whole face, which is necessary for new 3rd generation coding schemes like semantic or 3-D model based coding [9].

2 Quadtree image segmentation

A quadtree represents a digital image of a square region by hierarchically partitioning the region into quadrants and subquadrants until all subquadrants are uniform with respect to image value. This approach to image representation was first proposed by Klinger [5] and is very popular in the field of pattern recognition, image analysis and image coding [6,7,8]. As partitioning condition the tolerated variation of brightness within a single region is very common. A very important property of the quadtree image segmentation algorithm is the preservation of edges and detailed stuctures on the one hand and the merger of homogeneous areas to larger regions on the other hand. Especially this adaptive property predestines the quadtree image segmentation as good preprocessing candidate for the VBSVQ. The sizes of the finally created square regions are dependent on their positions in the quadtree. A region at tree–level l has the size of $2^l * 2^l$ pels. The quadtree segmentation algorithm is implemented in the programming language C to create training data for the image segmentation neural networks.

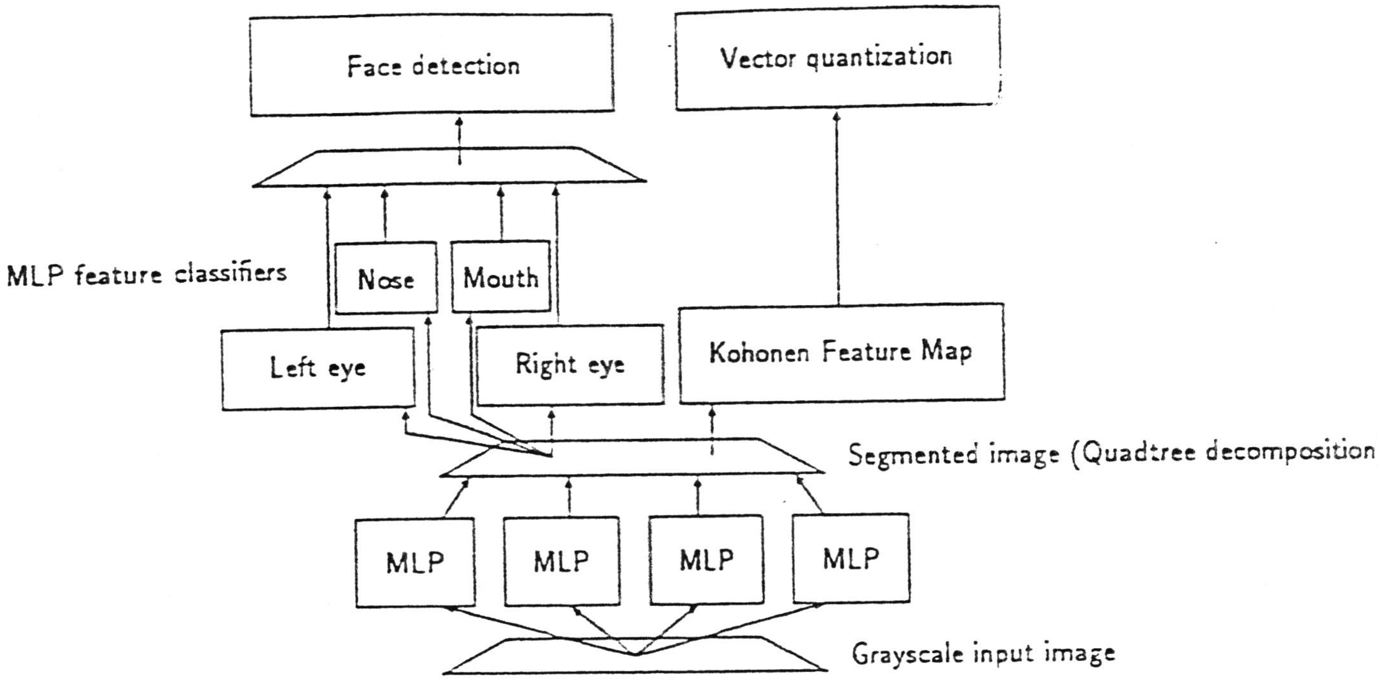

Figure 1: Neural Image Segmentation and Coding System

3 Neural net architectures

Corresponding to Figure 1 a first group of MLPs is implemented to do the segmentation step from a grayscale picture to its quadtree representation. Each MLP is responsible for a certain region block size. The first MLP investigates 2*2 pels image areas and decides if this region is homogeneous referring to its gray-scale distribution or not. The next one investigates an 4*4 pels large region, etc. Each neural network is a fully connected feed forward three layer perceptron. The MLPs are trained with the standard error backpropagation algorithm [11]. The Inputlayers of the MLPs have the same size as the investigated input image region (2*2,4*4,8*8,16*16). The image values (0..255) are scaled between -0.5 and 0.5. Each MLP creates a segmented image with only one block size. All output images together form a segmented image equivalent to its quadtree representation.
As shown in Figure 1 a Kohonen Feature Map is used to solve the vector quantization problem. This type of single layer, laterally connected neural net was introduced by T. Kohonen [10] and has the ability to realize a spatial organized clustering. This property allows to create codebooks in an initial learning step. The sizes of the individual codebook-blocks are similar to the region block sizes in the segmented image (2*2,...,16*16). The final codebooks are stored in the map and are used to assign the best matching codebook-block to the corresponding input picture block in a following recall step. Another posssible evaluation of the segmented image is the extraction of important facial features. Therefore another set of three layer perceptrons is implemented to search for eyes, nose and mouth. These MLPs are trained witch the standard error backpropagation learning law, too. Only the smallest block size (2*2 blocks) representation of the investigated image, containig the important edge information, is used for this operation because the spatial distribution of these blocks is sufficient for the recognition of facial features. Each MLP is trained with data of one feature, e.g left eyes, right eyes etc. The feature recognition of facial features in a binary image representation (only 2*2 blocks) seems to be a good application for the BACCHUS VLSI binary associative memory developed at DUT [1,2,3].

4 Simulation results

Figure 2 shows simulation results of the segmentation and VBSVQ processes. Starting with the 256*256*8 bits picture CLAIRE.LUM (a) Figure 2(b) is the result of a quadtree segmentation process done by four individually trained MLPs. Each MLP is responsible for the generation of 16*16, 8*8, 4*4 and 2*2 pels image blocks. Figure 2 (c–h) is a step-by-step demonstration of the VBSVQ. The segmented image (d) describes the location of the different sized image blocks. Each block is assigned to the best matching codebook-block. The generation of the different sized codebook-blocks and the assignment of the best matching blocks is done by a Kohonen Feature Map. The compression ratio of the shown 128*128*8 bits Picture is 14:1, the SNR is about 31 dB. The VBSVQ applied to the 256*256*8 bits Picture CLAIRE.LUM realizes a compression ratio of about 36:1 preserving the same good visual image quality.

5 Conclusion and Future Aspects

A neural image segmentation system has been presented. The segmented image is equivalent to its quadtree representation and serves as starting point for image coding and image analysis. The adaptive VBSVQ procedure offers increased compression rates *and* decreased image errors in comparison with the fixed block size VQ. It has to be mentioned that this system realizes *intraframe* coding, but an implemetation of the presented system in a hybrid coder similar to the MPEG or CCITT H.261 coder is possible. Future work will concentrate on this extension to interframe coding, coding schemes based on face detection and the application of the available binary associative memory BACCHUS in the field of face detection.

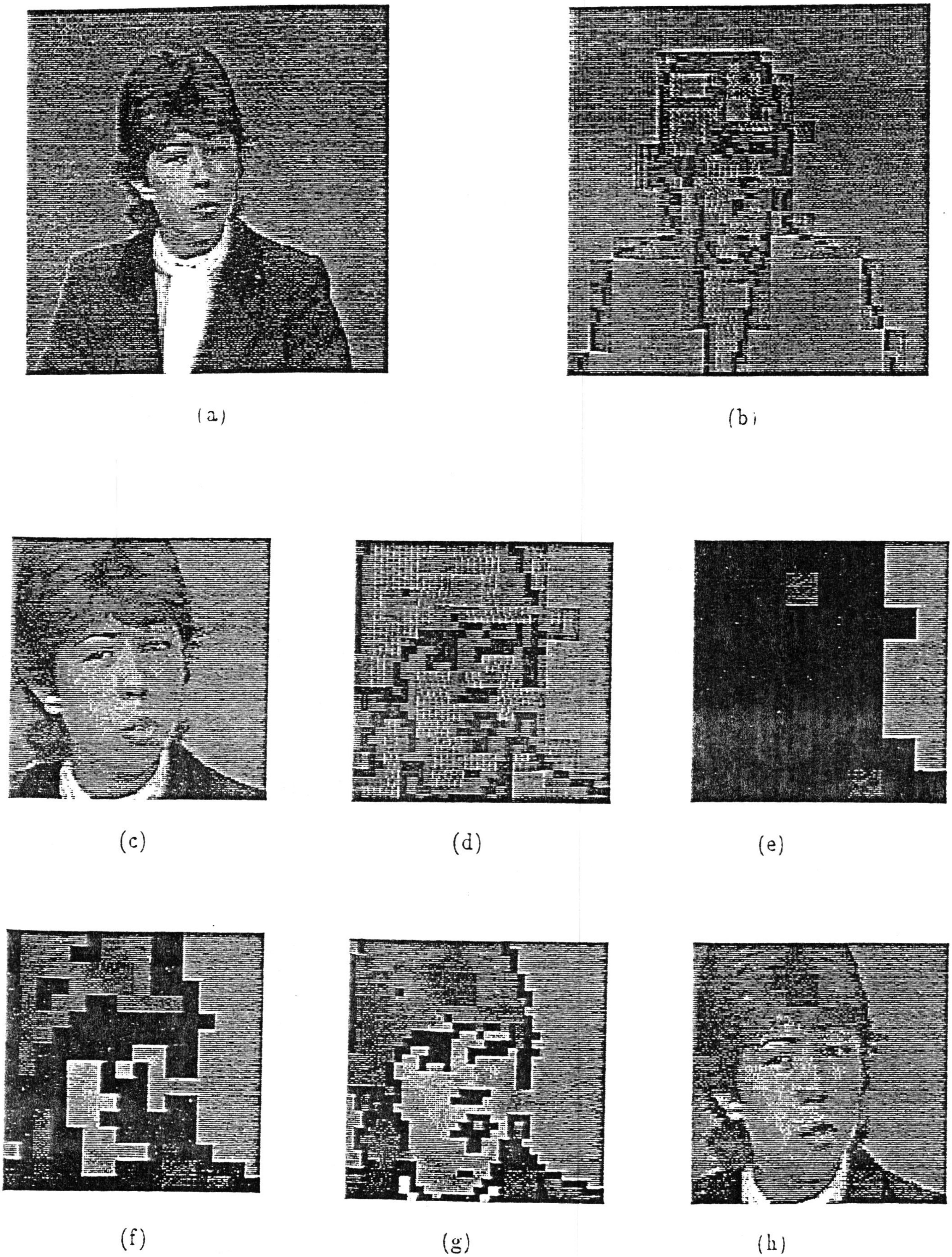

Figure 2: Simulation results

VOLUME II
OF III

IJCNN

INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS

Beijing, China
November 3–6, 1992

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

ISBN 7-5053-1906-X/TN.573
VOL I pp. 631-635