FANN–BASED VIDEO CHROMINANCE SUBSAMPLING

Adriana Dumitraş and Faouzi Kossentini

Department of Electrical and Computer Engineering,
University of British Columbia,
2356 Main Mall, Vancouver BC V6T 1Z4, Canada,
Fax: 604–822-5949, Email: adrianad@ee.ubc.ca, faouzi@ee.ubc.ca

ABSTRACT

In this paper, we present a video chrominance subsampling method using feedforward neural networks. Experimental results show that our method outperforms spatial subsampling obtained via lowpass filtering and decimation both objectively and subjectively. Other advantages of our algorithm are computational efficiency and low memory requirements. Moreover, no pre– or post–processing is required by our method.

1. INTRODUCTION

In color video coding, chrominance frames are usually subsampled, while the luminance component is not changed. This is possible due to the lower sensitivity of the human visual system to color information, as compared to the sensitivity to the luminance information [1], [2]. Prior to downsampling, signal conditioning in the form of lowpass filtering is commonly performed [3]. When applied to color video sequences, traditional lowpass filtering produces visible artifacts in the reconstructed video sequence. In order to minimize such artifacts, chrominance subsampling using feedforward artificial neural networks (FANNs) is proposed. We show that good objective/subjective results can be obtained with our pattern matching algorithm. The rest of the paper is organized as follows. The next section provides a brief description of our algorithm. Experimental results and conclusions are given in the last two sections.

2. FANN–BASED SUBSAMPLING ALGORITHM

Let the feedforward neural network (FANN) model be an M–H–N multilayer perceptron, i.e. with M input, H hidden and N output nodes [4] and let $N = 1$. The input image is stored as a matrix of samples, each matrix element representing a pixel value. A $W \times W$ window slides over the entire image, producing $M \times 1$ input vectors. For each input window (pattern) ξ, the FANN computes weighted sums of input values, passes them as arguments to nonlinear activation functions and outputs the actual value $y(\xi)$. During supervised training, the actual output value is compared with the desired output $d(\xi)$, the error $e(\xi)$ is computed and the FANN parameters are adjusted so that the chosen cost function (usually the sum of all squared $e(\xi)$'s) is minimized. If $\sqrt{M/N}$ is an integer greater than one, the FANN multilayer perceptron structure inherently performs subsampling. However, computing the error $e(\xi)$ for each input pattern ξ as mentioned above does not take into account the local characteristics (e.g. edges) inside of the current window. Therefore, some artifacts tend to appear in the reproduced frames. In order to avoid artifacts, we propose a supervised strategy to select the desired output value for the current input window. In the following, we summarize the proposed algorithm.

First, we compute the actual output value $y(\xi)$ for each input pattern ξ. Second, we compute the medians of all possible three–pixel combinations given by the intersection between the diamond (cross) shaped pattern shown in Figure 1, sliding over the search window, and the input window. We compare these median values with the fourth pixel in the window. Then, the pixel which is the closest to the median of the other three pixels is selected as the desired output value for the current input window. Fourth, we compute the global error at the end of one epoch[1] as

$$C(\mathbf{w}) = \frac{1}{2P} \sum_{\xi=1}^{P} \left[d(\xi) - y(\xi)\right]^2$$

where P is the total number of input patterns and $\mathbf{w}$ is the FANN weight (parameter) vector. Finally, we adjust all the weights according to a quasi–Newton rule. In this work, we have used the Levenberg–Marquardt approximation of the inverse Hessian matrix and a learning rate as given in [4]. The steps of the algorithm are repeated until the value of the error drops below a selected threshold or, until a predefined number of epochs is exceeded. Then, the FANN parameters are saved and used during the testing step. More details on

This work was supported by the Natural Sciences and Engineering Research Council of Canada under contract # 06P-0187668.

[1] Defined as one pass through the P–dimensional set of training patterns.

the above algorithm can be found in [5].

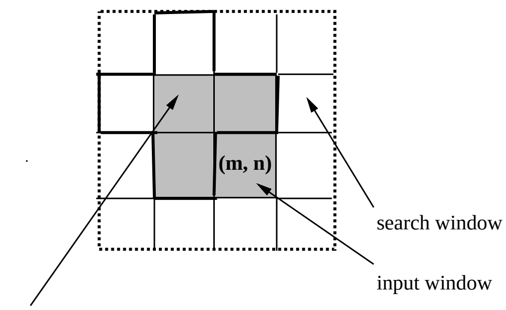

Figure 1: ONE OF THE PATTERNS GIVEN BY THE INTERSECTION BETWEEN THE DIAMOND (CROSS) SHAPED PATTERN (SLIDING OVER THE SEARCH WINDOW) AND THE INPUT WINDOW.

3. EXPERIMENTAL RESULTS

3.1. Implementation details

The block diagram of a chrominance subsampling system is shown in Figure 2. In the first experiment, two 4–2–1 FANN structures (one for the U frames and one for the V frames) were trained on a data set consisting of standard QCIF chrominance video frames from the sequences CLAIRE (frame 490), GRANDMA (frame 490), SALESMAN (frame 49), MISS AMERICA (frame 49) and SUZIE (frame 49). Each of the FANNs was tested on chrominance frames from the sequences TREVOR and MOTHER–AND–DAUGHTER. In the second experiment, two 4–8–1 FANN structures (one for the U frames and one for the V frames) were trained on a data set consisting of the first 10 chrominance frames from the standard sequences AKIYO, COASTGUARD and STUDENTS. The data set consisted of the first frame of each sequence, followed by the second frame of each sequence, etc. Each of the FANNs was tested on chrominance frames 100 to 250 from the sequence PARIS.

The size of the input window, the evaluation criteria and the other methods used in our comparisons, are important issues that need to be specified. In both experiments we have chosen a 2×2 window, moving one pixel to the right. The window was non–overlapping in the first experiment and overlapping the second experiment. We have chosen to evaluate the performance of our algorithm by (a) visual examination of the subsampled frames [6], (b) visual examination of the subsampled and cubic interpolated frames and (c) objective evaluation of the subsampled and cubic interpolated frames, based on peak signal–to–noise ratio (PSNR), mathematically given by

$$PSNR = 10\ \log_{10}\left(\frac{255^2}{MSE}\right), \text{ where}$$

$$MSE = \frac{1}{Q_1\,Q_2}\sum\left(x_{i,j}\ -\ \hat{x}_{i,j}\right)^2, \text{ and}$$

$\hat{x}_{i,j}$ are the pixel values in the $Q_1 \times Q_2$ reconstructed image. The results were compared to the those of the spatial lowpass filtering followed by quincunx subsampling (LPFS) method, for three 2–D filters: LPF1, which was designed via frequency sampling, LPF2, which was a separable finite response filter (FIR) designed using separable 2–D windows, and LPF3, which was a nonseparable FIR designed with 2–D windows. All of the filters have an order of 11 and a cutoff frequency equal to 0.5.

3.2. Performance and complexity of our algorithm

In the first experiment, gray level reproduction of the subsampled chrominance frame 75 of the TREVOR sequence shows that the FANN frames are the closest to the original chrominance frames, as illustrated in Figure 3. The dynamic range of the original U and V frames was $[79\ldots125]$ and $[122\ldots181]$, respectively. While the FANN maintains the dynamic ranges $[79\ldots122]$ and $[120\ldots176]$ for U and V, respectively, the LPF outputs values in the ranges $[37\ldots131]$ and $[49\ldots186]$ for U and V, respectively.

Then, we applied the FANN chrominance subsampling system to the U and V frames and encoded the QCIF test sequence MOTHER–AND–DAUGHTER (150 frames) using Telenor's H.263 video coder [7]. Both the PSNR values and visual quality of the obtained video sequence have been compared to the original ones, as well as to those given by the LPFS method. Several subjective evaluations of the decoded video sequence indicate that artifacts are present when chrominance frames have been lowpass filtered, while they are non–existent in the FANN subsampling (FANNS) case. Quantitative evaluation in terms of MSEs and PSNRs, for rates of 8 kbits/sec and 24 kbits/sec is given in Table 1. For rates between 4 kbits/sec and 32 kbits/sec, in steps of 2 kbits/sec, the results are displayed in Figure 4. The FANN gain is significant at low bit rates (1.9246 dB at 4 kbits/sec), and drops at higher rates (0.65 dB at 32 kbits/sec). The PSNR was obtained by:

$$PSNR = \frac{4\,PSNR(Y) + PSNR(U) + PSNR(V)}{6}.$$

In the second experiment, we applied the FANN chrominance subsampling system to the U and V CIF frames, then encoded the resulting YUV sequences at the rate of 24 kbits/pixel for the frames 100 to 250 of the sequence PARIS, using Telenor's H.263 video coder [7]. Next, we decoded the sequence and performed cubic interpolation. The PSNR values of the resulting frames w.r.t. the original chrominance frames are presented in Figure 5. On average, the FANNS outperforms the LPFS by 0.88 for the U frames and by 1.12

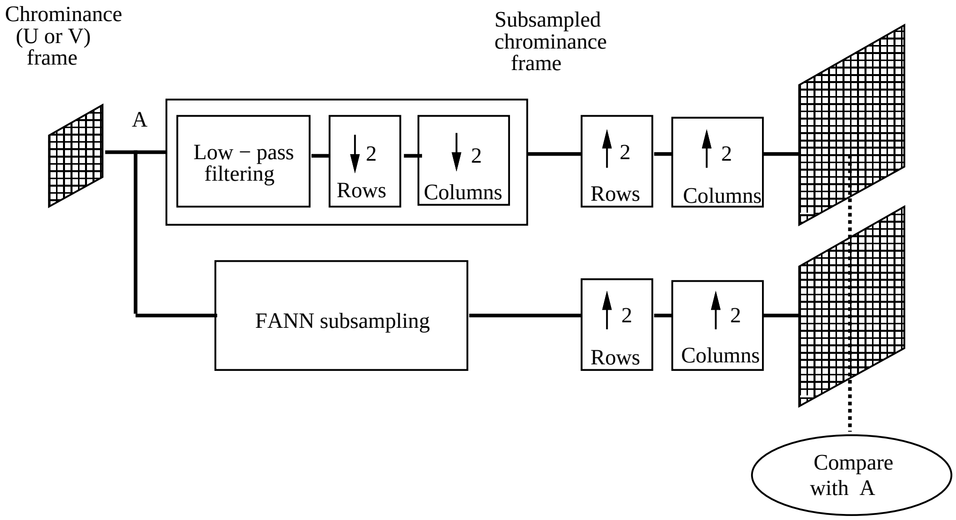

Figure 2: BLOCK DIAGRAM OF THE CHROMINANCE SUBSAMPLING SYSTEM.

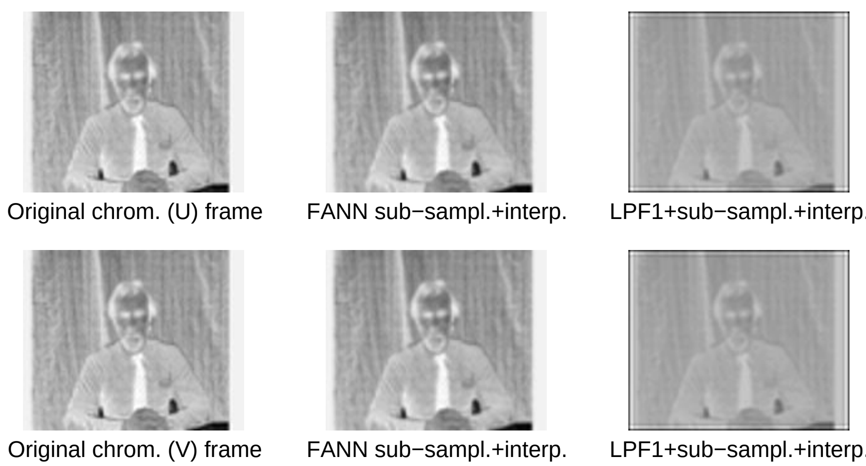

Figure 3: GRAY LEVEL REPRESENTATION OF THE CHROMINANCE FRAMES.

Table 1: PSNRs [dB] for different coding rates (8 kbits/sec and 24 kbits/sec), when using Telenor's H.263 video coder. Acronyms $FANNSI, LPFSI$ stand for FANN subsampling and LPF subsampling, both followed by cubic interpolation.

Method	8 kbits/sec			24 kbits/sec		
	Y	U	V	Y	U	V
FANNSI	30.878	37.214	37.381	34.295	39.768	39.610
LPF1SI	30.434	32.838	33.045	33.938	38.031	38.168

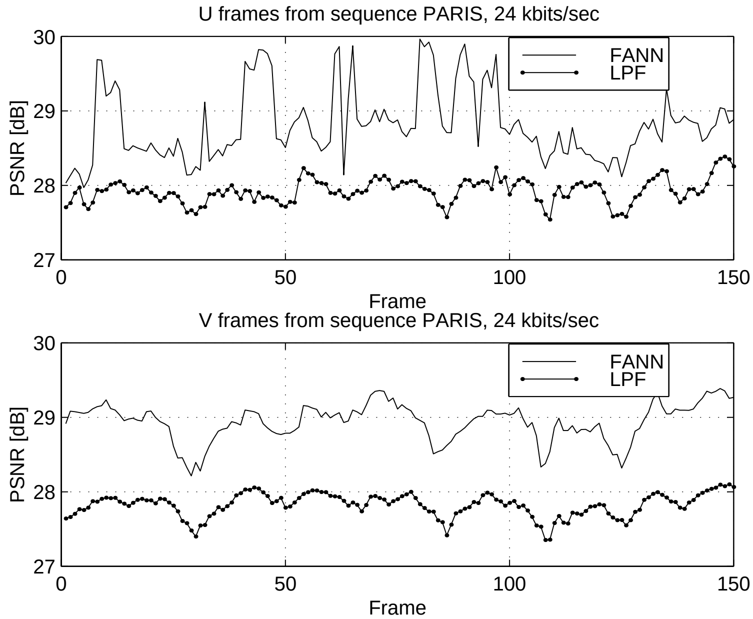

Figure 5: PEAK SIGNAL–TO–NOISE RATIO [DB] ON CHROMINANCE FRAMES 100 TO 250 OF SEQUENCE PARIS USING TELENOR'S H.263 VIDEO CODER AT 24 KBITS/SEC.

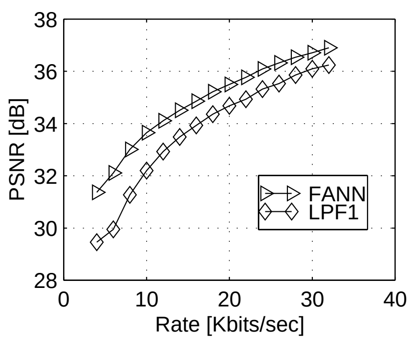

Figure 4: PEAK SIGNAL–TO–NOISE RATIO [DB] W.R.T. RATE IN LOW BIT RATE EXPERIMENTS USING TELENOR'S H.263 VIDEO CODER.

for the V frames. Subjective evaluation of the sequences indicates finer details in the FANNS sequences, as compared to the LPFS ones.

Finally, the speed of our FANN–based subsampler is evaluated by the test time on an UltraSparc 2 computer. The FANN requires only 0.12 seconds CPU time per QCIF frame, as compared to 5.5 times more, needed by the LPF method. Moreover, the designed FANNs feature memory requirements comparable to those of the LPFs, which makes our neural subsampling system well–suited for real–time applications.

4. CONCLUSIONS

In this paper, we have presented a subsampling algorithm and we have applied it to chrominance subsampling of video sequences. We have also presented experimental results for FANNS trained on single frames from different sequences (first experiment) and for FANNS trained on several frames from distinct sequences. Test results show that our method (a) leads to good objective and subjective video reproduction quality, (b) is computationally efficient, (c) has low memory requirements and (d) requires no pre– or post–processing. Our results outperform spatial subsampling obtained via the lowpass filtering and decimation method.

5. REFERENCES

[1] A. Murat Tekalp, *Digital Video Processing*, Prentice–Hall Inc., USA, 1995.

[2] K.R. Rao and J.J. Hwang, *Techniques and Standards for Image, Video and Audio Coding*, Prentice–Hall Inc., USA, 1996.

[3] Jelena Kovacevic and Martin Vetterli, "FCO sampling of digital video using perfect reconstruction filter banks," *IEEE Trans. on Image Processing*, vol. 2, no. 1, pp. 118–122, Jan. 1993.

[4] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*, John Wiley and Sons, USA, 1993.

[5] Adriana Dumitraş and Faouzi Kossentini, "Fast and high performance image subsampling using feedforward neural networks," Submitted to *IEEE Trans. on Image Processing*, 1997.

[6] R. Gonzalez and P. Wintz, *Digital Image Processing*, Addison–Wesley Publishing Company, Menlo Park, CA, 1987.

[7] Telenor Research, "TMN (H.263) encoder/decoder, version 2.0," *TMN (H.263) codec*, June 1996.