

Level Reassignment: A Technique for Bit-Rate Reduction

By B. PRASADA, F. W. MOUNTS, and A. N. NETRAVALI

(Manuscript received June 17, 1977)

Three techniques are described for reducing the entropy of a digitally coded monochrome picture signal by dynamically changing the input-output relationships of a single quantizer. This is done by reassigning the input levels of the quantizer to different representative output levels in such a way as to reduce the entropy of the quantized output, while keeping the visibility of the resultant quantization noise below a certain specified threshold. These techniques are simulated both in the pel and the transform domain for a simple differential pulse code modulation system and Hadamard transform coding system, respectively. It is found that they reduce the entropy by about 20 to 25 percent in the pel domain without noticeably sacrificing the picture quality. In the Hadamard transform domain, the entropy of the first transform coefficient can be decreased by about 23 percent with little change in picture quality.

I. INTRODUCTION

It is well known that the statistics of picture signals are nonstationary and that the required fidelity of reproduction demanded by the human eye varies from picture element (pel) to picture element. Consequently, for efficient digital representation of pictures, it is desirable to adapt coding strategies to those local properties of the picture signal which determine the visual sensitivity to quantization noise. In this study, we make use of the spatial masking properties of the human observer to adapt the quantization strategies for encoding the picture signal. We define spatial masking as the reduction in the ability of a person to visually discriminate amplitude errors which occur at or in the neighborhood of significant spatial changes in the luminance. To this end, we borrow, from our earlier work, measures of luminance spatial activity in both the transform^{1,2} and the pel³ domain, and the relationships (called the visibility functions) between the amplitude accuracy and

these measures of spatial detail. We use these visibility functions to change dynamically the input-output mapping of a single quantizer to reduce the bit rates. This is done by reassigning the input of the quantizer to a different representative level than normal in such a way as to reduce the entropy of the quantized output, while keeping the visibility of quantization noise below a certain threshold. We demonstrate three different algorithms for level reassignment and evaluate their potential by measuring the entropy of the quantized output for a given picture quality. It is worth pointing out that there are other methods of adapting the quantizer characteristics. Some of these are discussed in Refs. 3, 5, and 6.

1.1 Summary of the approach

To illustrate the coding strategy, let e_{in} and e_{out} be the input and output to a quantizer, whose decision levels are denoted by X_i , $i = 1, \dots, N + 1$ and representative levels by Y_i , $i = 1, \dots, N$. The quantizer input-output mapping in the absence of adaptation is:

$$\text{If } X_i < e_{in} \leq X_{i+1}, \text{ then } e_{out} = Y_i, \quad i = 1, \dots, N \quad (1)$$

In the adaptive quantization, instead of representing the input e_{in} as Y_i , we change the representation to Y_j ($j \neq i$) provided the visibility of the quantization noise ($e_{in} - Y_j$) is below a given threshold. The changed representation Y_j is chosen, in general, to decrease the entropy* of the quantizer output. This is done in several different ways. In one case, we increase the frequency of occurrence of the inner levels† of the DPCM quantizer. Here we make use of the fact that the frequency of occurrence of the inner levels of a DPCM quantizer is generally high, and, therefore, making it higher decreases the entropy. In another case, for alternate levels Y_i , Y_j is taken to be Y_{i-1} or Y_{i+1} , depending on the noise visibility. This has the effect of decreasing the frequency of occurrence of alternate quantizer levels, and increasing the frequency of occurrence of the other levels, and thus reducing the entropy. We refer to such a change in the use of the quantizer levels as level reassignment. Spatial detail surrounding the pel and the corresponding visibility function are used in determining whether quantization noise ($e_{in} - Y_j$) is visible. We have studied three different algorithms which differ only in detail from the approach outlined above. They are described in Section III.

* We define the entropy of the quantizer output as

$$\epsilon = - \sum_{i=1}^N \frac{n_i}{\sum_{j=1}^N n_j} \log_2 \left[\frac{n_i}{\sum_{j=1}^N n_j} \right]$$

where n_i is the frequency of occurrence of level Y_i .

† Inner levels are those which are close to level zero.

1.2 Summary of results

We simulated the level reassignment algorithms on a digital computer. The efficiency of an algorithm was measured by evaluating the difference in entropy between a nonadaptively coded picture and an adaptively coded (using level reassignment) picture having the same quality. In the pel domain, reassigning all the quantizer levels to the lowest level permitted by the quantization noise visibility leads to degradation at edges in the form of slope overload and, therefore, allows only about 9 percent reduction in entropy before picture quality is degraded. To avoid degradation of these edges, only a few inner levels (three inner levels on both sides of zero of a 15-level quantizer) were reassigned. This allows a higher threshold for the quantization noise before it is visible and leads to about 14 percent decrease in entropy. Alternate reassignment, in which certain alternate levels are reassigned to one of their two surrounding levels depending on the visibility of quantization noise, also allows a higher threshold for quantization noise and results in a 25 percent decrease of entropy. Delayed reassignment, in which reassignment is done only if the reassigning of the present sample to a lower level does not "adversely" affect the coding of the next sample, gives a 22 percent decrease in entropy. Thus alternate reassignment and delayed reassignment appear promising.

In the transform domain, we considered differential PCM coding of the first Hadamard coefficient. Using alternate reassignment, it was possible to reduce the entropy of the coded first transform coefficient by about 23 percent. We also give results on different methods of evaluating the visibility of the quantization noise and the effect of changing the picture content.

II. MASKING AND VISIBILITY FUNCTIONS³

We have constructed a simple measure of luminance spatial activity. We call this measure the masking function. In the pel domain, the masking function at a pel is taken as the weighted sum of the luminance slopes at the pel under consideration and at the neighboring pels. Thus, at point (i, j) the two-dimensional masking function $M_{i,j}$, using a 3×3 neighborhood of slopes is given by

$$M_{i,j} = \sum_{n=i-1}^{i+1} \sum_{t=j-1}^{j+1} \alpha \| (n,t) - (i,j) \|^{-1/2} \cdot [|m_{n,t}^H| + |m_{n,t}^V|] \quad (2)$$

where $\| (n,t) - (i,j) \|$ is the Euclidean distance between points (n,t) and (i,j) normalized by the distance between horizontally adjacent pels; $m_{n,t}^H$, $m_{n,t}^V$ are the horizontal and vertical slopes of the image intensity at point (n,t) ; α is a constant taken to be 0.35 based on the tests given in Ref. 3.

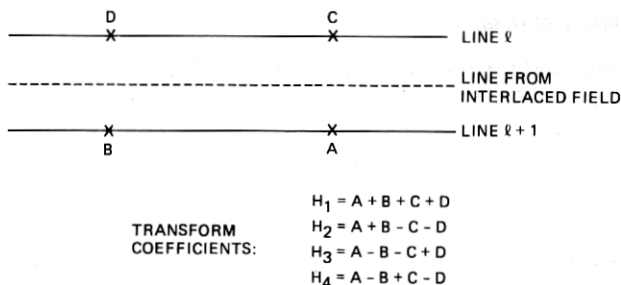


Fig. 1—Definition of Hadamard coefficients. A, B, C, D are the pel positions, and H_1, H_2, H_3, H_4 are the Hadamard coefficients.

In this framework, many masking functions are possible and some of these are discussed in Ref. 3.

In the Hadamard transform domain^{1,2} using a 2×2 transform of a block of pels in the same field defined as in Fig. 1, the measure of spatial activity in a block is taken as

$$H = \max(|H_2|, |H_4|) \quad (3)$$

It is clear from Fig. 1 that H is the maximum magnitude of the average line or element difference of pels within the block.

We hypothesized that the masking of the quantization noise is related to the spatial detail as measured by the masking functions (M or H). The precise relationship was obtained through subjective tests.* In these tests, the test picture was obtained by adding varying amounts of noise (to simulate the quantization noise) to all pels (or blocks in the case of transform domain) where the measure of spatial detail (M or H) has a given value. Such a picture was then compared in an A/B test with an unimpaired picture to which the subjects added controlled amounts of white noise everywhere in the picture until they found both pictures to be subjectively equivalent.

Visibility functions were derived from such subjective equivalence. They are shown in Fig. 2. Figure 2a shows the visibility of noise added to a pel as a function of masking function M , at that pel, whereas Fig. 2b shows the visibility of noise added to Hadamard transform coefficient H_1 as a function of $\max(|H_2|, |H_4|)$. Both these visibility functions decrease with respect to their argument, implying that at higher values of spatial detail the visibility of quantization noise is lower. Details of the subjective tests and visibility functions are given in Refs. 1-3. From this work we know that the visibility of noise at a point is proportional to the power of noise and the proportionality constant is given by the value of

* These tests are similar to the ones described by Candy and Bosworth.⁴

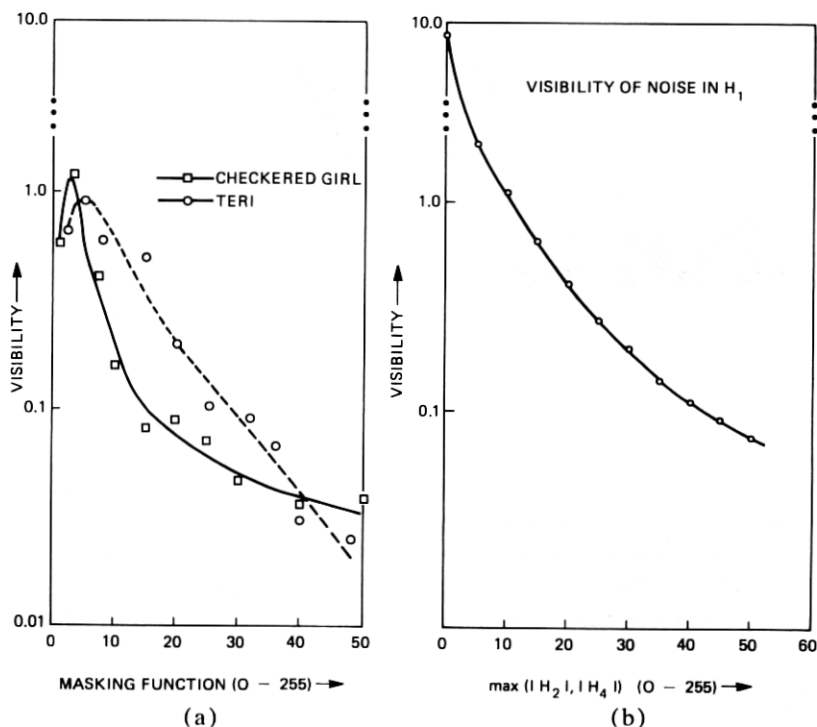


Fig. 2—(a) Visibility function for Checkered Girl and Teri using a two-dimensional one neighbor masking function. (b) Visibility function for noise added to H_1 as a function of $\max(|H_2|, |H_4|)$.

the visibility function evaluated using the spatial detail at that pel. Thus visibility functions allow us to judge the visibility of the quantization noise.

III. DETAILS OF THE TECHNIQUES

The entropy-reducing techniques described briefly in Section I have been simulated on a computer. In this section we give details of the technique as well as the results of the simulation. We have taken two different head and shoulders views (Fig. 3a, called Checkered Girl, Fig. 3b, called Teri) and processed them using our algorithms. We are dealing with a still picture, but since the quantization noise in a real television system varies from frame to frame even with still pictures, we coded three frames of the same picture to which a small amount of random noise (to simulate camera noise) was added. Due to the randomness of this noise from frame to frame, the noise appeared to move when the three frames were displayed in a repetitive sequence.



(a)



(b)

Fig. 3—Original pictures of (a) Checkered Girl and (b) Teri. These are 256×256 arrays with signal having a bandwidth of 1 MHz; each sample is linearly quantized with 8-bit PCM.

In the pel domain, we started with a DPCM coding system which had a 15-level quantizer optimized for minimum mean square quantization error. Using this quantizer, the coded pictures had noticeable defects, but were considered to be of acceptable quality by the authors. The entropy of the quantized output for Checkered Girl is 2.93 bits/pel and for Teri 2.64 bits/pel. In our experiments, the parameters of the coder were varied to decrease the entropy while keeping the quality of the coded picture similar to that of the 15-level DPCM coded picture. Thus the efficiency of the adaptive techniques described in the next section was judged by the difference between the entropy of the output of the adaptive coder and the 2.93 bits/pel (for Checkered Girl) and 2.64 bits/pel (for Teri) required for the nonadaptive coder. In the transform domain, we applied alternate reassignment to the DPCM coding of the first Hadamard coefficient using the first Hadamard coefficient of the horizontally adjacent block as the predictor. We used a 37-level quantizer which resulted in a picture of acceptable quality as judged by the authors and entropy of 3.75 bits/block for Checkered Girl and 3.76 bits/block for Teri for coding of H_1 .

3.1 Reassignment to the lowest level

The 15-level DPCM quantizer characteristic that was used in our simulations is shown in Fig. 4. The level numbers are also marked in the

	0	3	8	15	24	33	42	58
LEVEL	0 1	2 5	6 11	12 19	20 28	29 37	38 49	50 255
NO.	0	1	2	3	4	5	6	7

Fig. 4—A 15-level symmetric quantizer used for DPCM processing in the pel domain. The symbol

$$\begin{array}{c} z \\ \overline{x \quad y} \end{array}$$

means that inputs between x and y including x and y are represented by z .

Table I — Change in quantizer level distribution from nonadaptive to adaptive coder which reassigns all levels to the lowest possible level for Checkered Girl. Only non-negative levels are shown.

		ith level of adaptive coder							
		0	1	2	3	4	5	6	7
Total pel count per level		26566	10624	3029	1874	1206	678	363	590
jth level of nonadaptive coder	0	16825	16825	0	0	0	0	0	0
	1	14465	5465	9000	0	0	0	0	0
	2	6161	3424	1286	1451	0	0	0	0
	3	3307	532	302	1474	999	0	0	0
	4	1804	180	22	84	817	701	0	0
	5	965	78	10	13	43	480	341	0
	6	663	27	3	5	12	21	325	270
	7	740	35	1	2	3	4	12	93
									590

figure. Let the decision levels for the i th level be X_i , X_{i+1} and the representative levels be Y_i . The algorithm for positive i proceeds as follows:

If e_{in} is the input to the quantizer and $X_i < e_{in} \leq X_{i+1}$, i.e., without adaptation e_{in} would be quantized as Y_i , then,

(i) Evaluate the quantization error in representing e_{in} by the next lower level, i.e., error = $e_{in} - Y_{i-1}$.

(ii) Compute the visibility of this error by

$$e_{vis} = |e_{in} - Y_{i-1}|^\gamma \cdot f(M) \quad (4)$$

where M is the value of the masking function at the pel, $f(\cdot)$ is the visibility function, and γ is a constant.

(iii) If the visible error is less than a preassigned threshold then the input e_{in} can be represented as level Y_{i-1} . The above three steps are then repeated until the lowest level, at which either the visible error is above the threshold or the zero level is reached.

Similar steps can be followed when i is negative. Thus the lowest level (in magnitude) to which the input can be reassigned such that the visibility of the quantization is below a given threshold is taken as its representation. In the simulations γ was set to 1, and the threshold was varied until the lowest entropy was obtained for our picture quality.

The results of this simulation, for Checkered Girl, are shown in Table I. In this table the (j, i) th entry is the number of pels which would be quantized by $\pm j$ th level in a nonadaptive coder but are quantized by $\pm i$ th level in the adaptive coder. The first column gives the total number of pels quantized by a level in the nonadaptive coder and the first row gives the total number of pels quantized by a level in the adaptive coder. Thus,

Table II — Change in quantizer level distribution from nonadaptive to adaptive coder which reassigns the inner three levels to the lowest possible level for Checkered Girl. Only nonnegative levels are shown.

		ith level of adaptive coder							
		0	1	2	3	4	5	6	7
Total pel count per level		29526	10190	2453	460	1663	827	611	685
jth level of nonadaptive coder	0	16950	16950	0	0	0	0	0	0
	1	14533	5697	8836	0	0	0	0	0
	2	7229	5483	605	1141	0	0	0	0
	3	3917	1396	749	1312	460	0	0	0
	4	1663	0	0	0	0	1663	0	0
	5	827	0	0	0	0	0	827	0
	6	611	0	0	0	0	0	0	611
	7	685	0	0	0	0	0	0	0

levels ± 2 had 6161 pels in the nonadaptive coder out of which 3424 went to level 0, 1286 went to level ± 1 , and 1451 remained in level ± 1 , in the adaptive coder. Clearly the distribution of pels using different quantizer levels in the adaptive coder is more peaked than the corresponding distribution in the nonadaptive coder. This causes a decrease in the entropy of the coder output. The decrease in the entropy without any noticeable change of picture quality is about 0.27 bits/pel which is approximately 9.2 percent.

One of the reasons for such a small decrease of entropy is the feedback process inherent in a DPCM coder. If, in quantization of a pel, the adaptive quantizer changes its occupancy from level $\pm j$ to $\pm k$ ($k < j$), then the differential signal to be quantized for the next pel generally increases in magnitude, and there is a good chance that it will occupy a higher quantization level than in the case of the nonadaptive coder. This, to some extent, hampers our original objective of making the distribution of pels in quantizer levels highly peaked. This feedback affects the very high levels more severely. Quantizing a "higher level sample" by a lower level, results in spreading of the picture edges, which is subjectively very annoying.

To obviate this, and knowing that the higher quantization levels have little effect on the entropy, we used the above algorithm to reassign only the inner ± 1 , ± 2 , and ± 3 levels. The results of this simulation are shown in Table II. The 0th level contains more pels than before. This reduces the entropy to 2.51 bits/pel which is a reduction of about 14 percent without any noticeable sacrifice of picture quality.

Using the inner ± 1 , ± 2 , ± 3 level reassignment algorithm, we studied the effects of the variation of γ in eq. (4). Small γ results in good repro-

duction of flat areas, but results in severe edge effects similar to the edge busyness. A value of γ larger than 2, on the other hand, results in better edge reproduction but shows granular noise in flat areas. We used three different values of γ , namely 0.5, 1.0, and 2.0, and varied the threshold T such that for each case we got the lowest entropy without changing the picture quality. The entropy for $\gamma = 0.5$ was 2.62 bits/pel, for $\gamma = 1.0$, it was 2.51 bits/pel and for $\gamma = 2.0$, it was 2.44 bits/pel. Thus, $\gamma = 2.0$ was clearly superior to either 0.5 or 1.0. We used $\gamma = 2.0$ for all our subsequent simulations.

3.1.1 Alternate reassignment

Alternate reassignment changes alternate levels (positive or negative) to one of the two adjacent levels depending on the quantization error visibility. This has the effect of reducing the level occupancy of alternate levels, thereby reducing the entropy. Also the highest level (positive or negative) was not reassigned for better edge reproduction. In our simulation of this scheme only alternate levels i.e., $\pm 1, \pm 3, \pm 5$ were reassigned. Level 1, for example, was reassigned to either level 0 or level 2 depending upon both of the following conditions:

(i) Reassign to level 0 if

$$|e_{in} - Y_0|^\gamma \leq |e_{in} - Y_2|^\gamma \quad (5)$$

otherwise reassign to level 2; and

(ii)

$$|e_{in} - Y_k|^\gamma f(M) \leq T \quad (6)$$

$k = 0$ or 2 , whichever satisfies condition 1.

Level 1 was not reassigned if both the above conditions were not met. Here again, both the positive and negative levels were reassigned similarly. This resulted in further improvement in the entropy of coded output. We tried alternate reassignment on levels ± 1 and ± 3 , and ± 5 with results as shown in Table III. Note that as expected, the number of pels in levels $\pm 1, \pm 3$, and ± 5 have decreased significantly. The entropy of the quantized output for this simulation was 2.19 bits/pel, which is a decrease of about 25 percent.

3.1.2 Delayed reassignment

Another way to counter the feedback in the DPCM coder is to wait until the next sample for the coding of the present sample and reassign the present sample only if the next sample is not "adversely" affected by reassigning the present sample. The reassignment of the k th sample

Table III — Change in quantizer level distribution from nonadaptive to adaptive coder which does alternate reassignment of levels ± 1 , ± 3 , ± 5 , for Checkered Girl. Only nonnegative levels are shown.

		ith level of adaptive coder							
		0	1	2	3	4	5	6	7
Total pel count per level		34802	1845	7868	90	2642	1	852	621
jth level of nonadaptive coder	0	16274	16274	0	0	0	0	0	0
	1	22192	18528	1845	1819	0	0	0	0
	2	4322	0	0	4322	0	0	0	0
	3	2561	0	0	1727	90	744	0	0
	4	1408	0	0	0	1408	0	0	0
	5	799	0	0	0	490	1	308	0
	6	544	0	0	0	0	0	544	0
	7	621	0	0	0	0	0	0	621

to the lowest possible level is made only

(i) If the visibility of error, i.e., $|e_{in} - Y_j|^\gamma \cdot f(M)$, is less than a threshold T_1 ; and

(ii) If the prediction error for the next sample does not change by more than " T_2 ", a specified value.

In our simulation, we set the threshold T_1 at a high value and varied T_2 between 1 and 15. $T_2 = 5$, which implies that reassigning the present sample should not change the prediction error of the next sample by more than 5, gave the best results. The results of this simulation for

Table IV — Change in quantizer level distribution from nonadaptive to adaptive coder which does delayed reassignment of all the levels for Checkered Girl. Only nonnegative levels are shown.

		ith level of adaptive coder							
		0	1	2	3	4	5	6	7
Total pel count per level		37659	5235	3181	1742	1209	710	504	637
jth level of nonadaptive coder	0	13946	13946	0	0	0	0	0	0
	1	23171	21219	1952	0	0	0	0	0
	2	7826	2020	3170	2636	0	0	0	0
	3	2454	347	97	446	15640	0	0	0
	4	1405	90	16	90	129	1080	0	0
	5	804	19	0	8	45	97	635	0
	6	591	9	0	1	4	29	70	478
	7	680	9	0	0	0	3	5	26

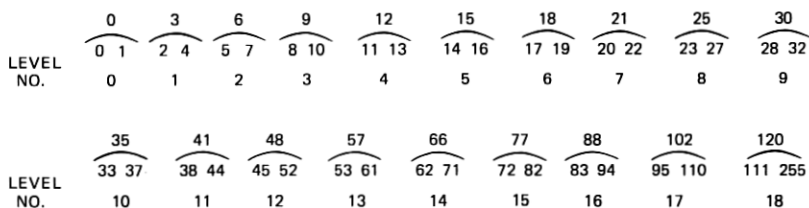


Fig. 5—A 37-level symmetric quantizer characteristic used for DPCM processing of the first Hadamard coefficient. (Only the positive portion is shown.)

Checkerboard Girl are shown in Table IV. It is seen that the highest level occupancy did not change significantly in the adaptive coder, but occupancy in the lower levels changed rather significantly. The resulting entropy was 2.28 bits/pel, a decrease of about 22 percent over the nonadaptive coder. It appears that delayed reassignment and alternate reassignment perform somewhat similarly in their bit-rate reduction capabilities.

3.2 Reassignment in the Hadamard domain

Alternate reassignment in the Hadamard domain was simulated for the first Hadamard coefficient " H_1 " of a 2×2 block. As mentioned in our earlier work,¹ we did DPCM coding of the first coefficient to make use of some of the statistical redundancy not utilized due to small block size. It was found that, when all the other coefficients, i.e., H_2 , H_3 , and H_4 , were not quantized, the 37-level quantizer whose characteristics are shown in Fig. 5, gave a satisfactory picture without any adaptation. Using this quantizer, alternate reassignment was simulated for levels ± 1 , ± 3 , ± 5 , ± 7 , ± 9 , ± 11 . The entropy of the quantized output was 2.90 bits/block for coding of H_1 , which amounts to a 23 percent decrease over that obtainable from the nonadaptive coder.

3.3 Sensitivity to picture variation

In order to study the sensitivity of the reassignment algorithms to picture variation, we coded Teri using all the above techniques. The visibility functions used were those obtained for Teri. The results are shown in Table V. It is clear from this table that the percentage decrease in entropy using reassignment is a little higher for Teri than for Checkerboard Girl. For example, using alternate reassignment there is about 30 percent decrease in entropy. We also coded Teri and Checkerboard Girl with each other's visibility functions and found relatively insignificant changes in entropy. However, the thresholds for the least entropy had to be different for the two pictures. Taking the lower value of the

Table V — Entropy results with level reassignment for Checkered Girl and Teri

Algorithm	Entropy (bits/pel)	
	Checkered Girl	Teri
Reassigning all levels to lowest level	2.66	2.31
Reassigning inner ± 3 levels to lowest level	2.44	2.23
Alternate reassignment on $\pm 1, \pm 3, \pm 5$ levels	2.19	1.87
Delayed reassignment	2.28	1.91
Hadamard transform coding, alternate* reassignment on $\pm 1, \pm 3, \pm 5, \pm 7, \pm 9, \pm 11$ levels	2.90	2.84

* Without level reassignment, the entropy of the coded H_1 coefficient was 3.77 bits/block for Checkered girl and 3.76 bits/block for Teri.

threshold and using it for both pictures increased the entropy for one of the pictures by about 5 percent.

IV. SUMMARY

We have described several coding algorithms to change dynamically the input-output relationships of a single quantizer in order to achieve bit-rate reduction. The algorithms reassign the input to the quantizer to a different representative level in such a way as to reduce the entropy of the output, while keeping the visibility of the quantization below a specified threshold. The techniques were demonstrated for DPCM coding systems both in the pel and the Hadamard transform domain. Although we do not discuss it here, it is possible to extend these techniques to PCM quantizers. The inherent feedback, which exists in a DPCM system, counters the reassignment strategy to some extent. Methods were devised to decrease the effect of feedback on the bit-rate reduction. It is possible to reduce the entropy by about 20 to 30 percent by our adaptive techniques, without changing the picture quality.

V. ACKNOWLEDGMENTS

Our work was greatly simplified by the use of the picture-processing computer system developed mainly by Jean-David Beyer and Ralph C. Brainard. We are grateful to Alan S. Kobran for programming assistance and to Barry G. Haskell who developed many of the general-purpose programs which we used.

REFERENCES

1. F. W. Mounts, A. N. Netravali, and B. Prasada, "Design of Quantizers for Real Time Hadamard Transform Coding of Pictures," B.S.T.J., 56, No. 1 (January 1977), pp. 21-48.
2. A. N. Netravali, B. Prasada, and F. W. Mounts, "Some Experiments in Adaptive and Predictive Hadamard Transform Coding of Pictures," B.S.T.J., 56, No. 8 (October 1977), pp. 1531-1547.

3. A. N. Netravali and B. Prasada, "Adaptive Quantization of Picture Signals Using Spatial Masking," *Proceedings of IEEE*, April 1977, pp. 536-548.
4. J. C. Candy and R. H. Bosworth, "Methods for Designing Differential Quantizers Based on Subjective Evaluations of Edge Busyness," *B.S.T.J.*, 51, No. 7 (September 1972), pp. 1495-1516.
5. E. R. Kretzmer, "Reduced-Alphabet Representation of Television Signals," *IRE Nat. Conv. Rec.*, 4, Part 4, 1956, pp. 140-147.
6. J. O. Limb, "Adaptive Encoding of Picture Signals," *Symposium on Picture Bandwidth Compression*, MIT, Cambridge, April 1969.

