EXHIBIT I

TO DECLARATION OF S. MERRILL WEISS IN SUPPORT OF PLAINTIFF ACACIA MEDIA TECHNOLOGIES CORPORATION'S MEMORANDUM OF POINTS AND AUTHORITIES IN OPPOSITION TO ROUND 3 DEFENDANTS' MOTION FOR SUMMARY JUDGMENT OF INVALIDITY UNDER 35 U.S.C. § 112 OF THE '992, '863, AND '702 PATENTS; AND SATELLITE DEFENDANTS' MOTION FOR SUMMARY JUDGMENT OF INVALIDITY OF THE '992, '863, AND '720 PATENTS

A Motion-Compensated Interframe Coding Scheme for Television Pictures

YUICHI NINOMIYA AND YOSHIMICHI OHTSUKA

Abstract—A motion-compensated prediction system for interframe coding is proposed. An interframe coding system is developed using this prediction method and second-order temporal prediction. Real-time hardware of reasonable circuit size using this coding system for monochrome TV pictures has been constructed. Sampling frequency and bit-rate of this system are 10.06 MHz and 1.6 bits/pel, respectively.

Assessments have been made of the picture quality of coded/decoded pictures from a variety of broadcast television programs. Quality was good for almost all such programs.

A subjective test of picture quality was performed under stringent conditions, using some of the most difficult scenes selected in earlier experiments. The lowest value for the quality of pictures coded/decoded at 1.6 bits/pel was 3.7 on a five-grade impairment scale.

I. INTRODUCTION

INTERFRAME coding systems offer a good solution to the problems of digital television signals. The most important merit of interframe coding is a very high coding efficiency for still pictures. Since human visual perception is most critical of still pictures, interframe coding is a good way to keep picture quality high.

However, interframe coding systems have a lower efficiency for moving pictures. Further, the sensitivity of human vision to image degradation is not much reduced during simple movement of the object such as occurs when "panning."

Most broadcast television programs include many moving scenes. It follows that conventional interframe coding systems are not appropriate to broadcast television coding with a high bit-reduction ratio.

Several proposals have been made for motion compensation by interframe prediction [1]. The basic idea of motion compensation is a very simple one. But it proved difficult to determine the motion vectors of the various parts of a television picture. The design of practical hardware was also difficult.

Several methods of determining the motion vector of an object in a television scene have been proposed. There are two basic methods; one of these is by pattern-matching (point-by-point correlation); and the other utilizes the relationship between the spatial differential signal and the temporal differential signal (Limb and Murphy's algorithm [2]). When this work was started, the authors felt that the matching method would result in easier design of first hardware because the required operations are only addition and comparation. Nevertheless, with a conventional matching system, the total amount of calculation is still large, if we want to construct a real-time system. So it could be said that the matching method is too complicated, especially if displacement needs to be defined with resolution finer than the sampled grid of the digital television picture. We propose an iterative method which was suggested by one of the authors [3] in order to decrease the number of calculation steps. For fine resolution, we also propose a simple detection method in which the motion vector is given as an integral multiple of the sampling distance.

Experiments have been performed on motion-compensated interframe coding [4], [5] systems. Many of these experimental systems are effective for certain simple types of motion, for example, "panning" by the TV camera. But in broadcast programs, motions are not always so simple. A system proposed by Netravali and Robbins [5] is effective for more complex motions. All these experimental results suggest that coding efficiency could be roughly doubled. Almost all work on motion compensation has been done by computer simulation, especially systems effective for more complex motions. The experiments described in this paper were performed using the local decoder output, and all experimental apparatus is real-time hardware.

Paper approved by the Editor for Data Communication Systems of the IEEE Communications Society for publication without oral presentation. Manuscript received September 15, 1980; revised September 30, 1981.

The authors are with the Advanced Broadcasting Research Division, Technical Research Laboratories of Nippon Hoso Kyokai, Tokyo, Japan.

II. METHOD OF DETERMINING THE MOTION VECTORS

A. Algorithm

The basis of the proposed algorithm is a block-by-block matching method. First, every television picture field is subdivided into blocks. We shall assume that displacement is constant within a single block.

We denote the level of the picture signal at the x_1th point of the x_2th line of the Nth field as

$$A^N(x_1, x_2) = A^N(x), \quad \text{where } (x_1, x_2) = x. \tag{1}$$

The measure of correlation is given by

$$D(y) = \sum_{x \in C} f(A^N(x) - A^{N-2}(S(y, x))$$

$$S(y, x) = x + y \tag{2}$$

where C is a subspace of the block, f is a measuring function, and S is a shift operator.

The motion vector is a vector $\hat{y}$ such that $D(\hat{y})$ is minimum. The function f should be simple enough so as to be easily cal-

0090-6778/82/0100-0201$00.75 © 1982 IEEE

Exhibit I Page 183

culated; for example

$$f_1(a) = T_t(|a|) \quad (3)$$

$$T_t(a) = \begin{cases} 1, & a > t \\ 0, & a \leqslant t \end{cases}$$

$$f_2(a) = u \cdot \log_2 (|a| + 1) \quad (4)$$

where T_t is a threshold function, u is a function which counts a fraction as a whole number, and $f_2(a)$ is the number of bits necessary to binary code the positive number equal to or less than $|a|$. We use f_2 in our experimental system.

To determine the motion vector we take $\hat{y}$ as follows.

$$\hat{y} \in \{\bar{y} \mid \forall y \in Y, \quad D(\bar{y}) \leqslant D(y)\}. \quad (5)$$

Here Y is a set of shift vectors from among which a search can be made for the minimum $D(y)$. We call such a set of vectors "a set of menu vectors" or "a menu"; the menu must include a large number of vectors. For example, in ordinary television programs, the speed exceeds 10 pels/frame (at 10.7 Msamples/s). It follows that the menu must include over 200 shift vectors and, consequently, the real-time detection of motion vectors is difficult with such a primitive method.

We therefore propose an iterative method [3]. The basic idea of this method is to divide the procedure for finding the minimum $D(y)$ into successive stages.

We denote the menus of the stages by $Y_1{}^0, \cdots, Y_n{}^0$. The numbers of elements in each menu are all of the same order. The mean difference between neighboring members of successive sets is decreased in accordance with the suffixes of the menus. This means that the larger the suffix, the more accurate is the determination of the motion vector, and at the same time the smaller the detection areas. For example

$$Y_i{}^0 = \{(y_1, y_2) | y_1, y_2 : 0, \pm 2^{n-i}, \pm 2^{n+1-i}\}, \quad i = 1 \cdots n$$

where (y_1, y_2) denotes a vector and $n = 4$.

The iterative estimation of $\hat{y}$ is as follows.

First, an initial approximation $\hat{y}_1$ of the motion vector is given by

$$\hat{y}_1 \in \{\bar{y}_1 \mid \forall y_1 \in Y_1{}^0, D(\bar{y}_1) \leqslant D(y_1)\}.$$

In the second stage, a second approximation $\hat{y}_2$ of the motion vector is obtained from the shifted menu Y_2 as follows.

$$\hat{y}_2 \in \{\bar{y}_2 \mid \forall y_2 \in Y_z, D(\bar{y}_2) \leqslant D(y_2)\}$$

where the shifted menu Y_2 is

$$Y_2 = \{y_2{}^0 + \hat{y}_1 \mid y_2{}^0 \in Y_2{}^0\}$$

and $y_2{}^0$ is an element of the $Y_2{}^0$ which is a fixed menu for the second stage.

We can estimate the motion-vector precisely by using this process iteratively. In the mth stage, $\hat{y}_m$ is given by

$$\hat{y}_m \in \{\bar{y}_m \mid \forall y_m \in Y_m, D(\bar{y}_m) \leqslant D(y_m)\} \quad (6)$$

where

$$Y_m = \{y_m{}^0 + \hat{y}_{m-1} \mid y_m{}^0 \in Y_m{}^0\}, \quad (m = 2, 3, \cdots, n).$$

If the distance between neighboring elements of the menu of the last stage $Y_n{}^0$ is sufficiently small, we can determine the motion vector as $\hat{y}_n$. If this procedure is to produce the correct motion vector, it is necessary that the function D should have a unique minimum. This is shown in the Appendix.

The same principle can be used to determine the motion vector with an accuracy greater than the sampling distance. For this purpose we must add further stages to those of (6). In these additional stages, fractional parts of the motion vector can be determined. As an example, to determine the motion vector with an accuracy of half a sampling interval, a stage is added, and at this stage D is given by

$$D(y) = \sum_{x \in C} f(A^N(x) - \tilde{S}(y, A^{N-2}(x)),$$

$$y = (y_1, y_2), \qquad y_1, y_2 \in \{-\tfrac{1}{2}, 0, \tfrac{1}{2}\},$$

$$\tilde{S}(y, A^{N-2}(x)) = \frac{A^{N-2}(x_1, x_2) + A^{N-2}(x_1 + 2y_1, x_2 + 2y_2)}{2}.$$

However, even for the method described above, hardware realization is somewhat difficult, as the calculation steps are still too complicated.

For actual television pictures, the number of stages needed can be reduced to only one. In television pictures, motion is usually temporally constant except at object boundaries, so we can use a simplified form of the iterative method as follows.

$$\hat{y} = \{\bar{y} \mid \forall y \in \bar{Y}, \ D(\bar{y}) \leqslant D(y)\} \quad (7)$$

$$Y = \{y + \hat{y}^{-1} \mid y \in Y^0\} \cup \phi \quad (8)$$

where ϕ is a zero vector and $\hat{y}^{-1}$ is the motion vector of the block in the same position of the preceding field. (Although, strictly speaking, the block in the preceding field is offset by one line from the current frame.) ϕ is necessary to reset the components of the shift vector to zero when the iteration moves into an incorrect region while searching for the motion vector. If the iteration moves into an incorrect region, values of all $D(y)$ except $D(\phi)$ become unusually large.

For such a system, the distribution of vectors in the menu Y^0 should be nonuniform, and the distance between neighboring vectors should be minimum at the center of Y^0. An example of such a menu is given in Fig. 1.

With this method, hardware realization is feasible, as described in the sequel. The motion-compensated prediction signal is given as the preceding frame signal shifted by the motion vector.

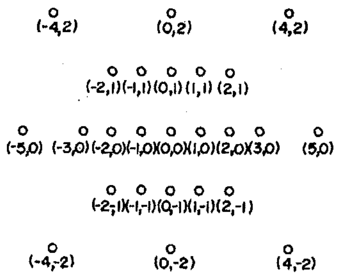


Fig. 1. "Menu" for determination of motion vector.

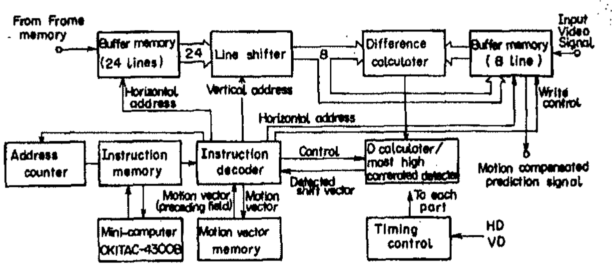


Fig. 2. Block diagram for motion vector detection and motion compensation.

B. Hardware Configuration

A block diagram of experimental hardware for motion-compensated predictor is given in Fig. 2. The preceding frame signal and current frame signal are stored in separate small buffer memories to detect the motion vectors. The vertical and horizontal position shifts are performed by the line shift and by the address control, respectively, of the buffer memory. The current frame buffer memory is also used as the output buffer for the motion-compensated signal.

The entire system of motion detection and compensation is sequentially controlled by instructions stored in the instruction memory. Control instructions are generated by the minicomputer and then loaded into the instruction memory. Such a configuration is appropriated for an experimental system, but for practical use, systems in which ROM's are used for the instruction memory, or in which microprocessors are used instead of a minicomputer, are preferable.

The logic circuits used in the experimental system were TTL's (mostly S-TTL's, including the LS-type).

III. CODING SCHEME

A. Outline of the Coding System

The coding system is basically block-wise variable word-length. The prediction modes include motion-compensated prediction and second-order temporal prediction needed for "fade" or "cross-fade" scene changes.

Sampling modes include normal sampling, mixed sampling, subsampling, and field-repeat. Mixed sampling is a mode in which subsampling is used only in "flat" regions of the pictures.

For regions where interframe prediction errors are high (for example, fast moving objects such as a man walking past the camera at just a short range), a rough intrafield DPCM scheme which has only three levels (for convenience, we call it ΔM) is used.

A complete block diagram is given in Fig. 3. The generation of information is controlled by controlling the coding threshold, the sampling mode and the quantizers.

B. Prediction

Motion-compensated prediction is the one used most often. Second-order temporal prediction is only used occasionally for "fade" scene changes.

1) Motion-Compensated Prediction: The following parameters were chosen for motion compensated prediction.

Block size: 8 lines × 16 pels.

Y^0 : as in Fig. 1.

C : subsampled to half (as shown in Fig. 4).

Measure function: f_2.

The use of motion-compensated prediction in an interframe coding system produces several problems. If the coding threshold is high, the "dirty window" noise moves with the motion vector. In a block which includes both a moving area and a stationary area, the motion of "dirty window" noise impairs stationary areas. To avoid this effect, we introduce a switching method as follows. The switches correspond to the S_4's in Fig. 3.

An uncompensated signal is used instead of the motion-compensated signal as the prediction signal at a picture element at a point x if

$$|A_M{}^{N-2}(x) - A_0{}^{N-2}(x)| \leq T_{M0} \tag{9}$$

and

$$|A^N(x) - A_0{}^{N-2}(x)| \leq T_0 \tag{10}$$

where A^N is an input signal level, $A_M{}^{N-2}$ and $A_0{}^{N-2}$ are compensated and uncompensated prediction signal levels, respectively, T_{M0} is a fixed threshold, and T_0 is the coding threshold. At the decoder, detection of the condition of (9) can be executed in the same way as in the coder. To detect the condition of (10), we can use the following condition

$$U(x) = 0$$

where $U(x)$ is the transmitted prediction error signal.

2) Fade Prediction: The configuration for fade prediction is very simple, as shown in Fig. 5 [8]. Frame-difference signals for each field are temporally smoothed and fed into the adder.

Motion-compensated prediction and fade prediction are switched fieldwise by S_1 in Fig. 3.

The switching algorithm is determined by experiment. If the prediction error in the fade prediction is less than 7/8 that of the motion-compensated prediction in three successive fields, the prediction is switched to fade prediction. Conversely, if this condition is not satisfied for five successive fields, it switches to motion compensation.

Exhibit I Page 185

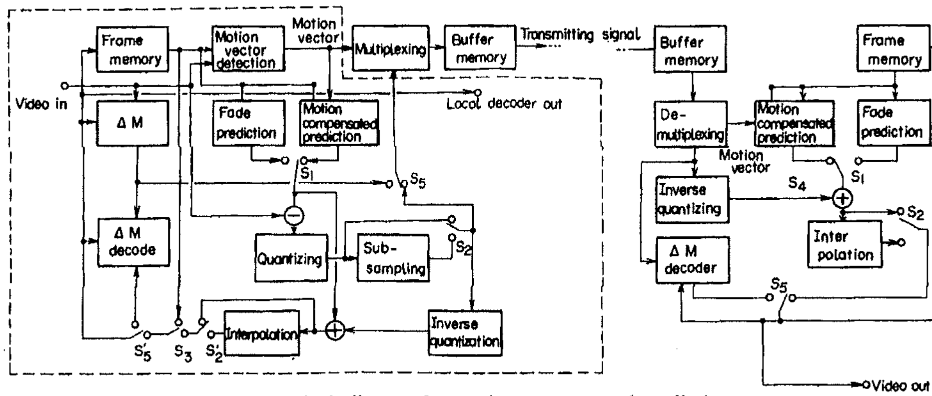


Fig. 3. Block diagram for motion compensated prediction.

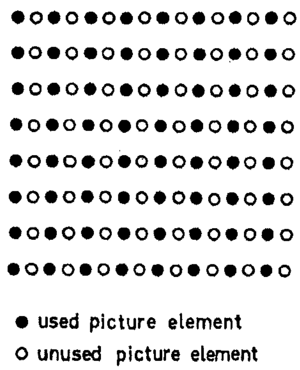


Fig. 4. Pattern of C.

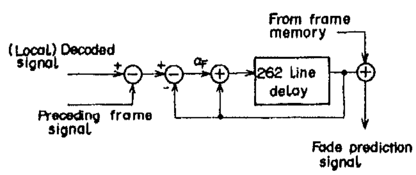


Fig. 5. Block diagram for fade prediction.

C. Transmission Format

We use a block-type variable word-length system for transmission. Transmission formats are shown in Fig. 6. The picture element flag is one if the frame prediction error signal to be sent is not zero. Bits are assigned for transmission of a prediction error signal only when the flag is one.

In each block, the word length used for transmission is constant, and the word length used is designated with 3 bits. The finest quantizer we used was nonlinear and of 7 bits. Where the block word length equals zero, the block flag is zero. There is a surplus position of word length determined by three bits used to indicate the use of ΔM as described below.

For the mixed mode, a extra bit is used to designate subsampling. In any mode, normal sampling and subsampling are switched blockwise.

D. Sampling Mode

We use four sampling modes: normal mode, mixed mode, subsampling mode, and field-repeat mode. These modes are switched fieldwise. The sampling frequency of the normal mode is 10.06 MHz ($f_H \times 640$; f_H is the horizontal frequency).

1) Subsampling Mode: For a rapidly moving picture, it is better to use subsampling than to use a very high coding threshold. This is because of the following:

a) the sensitivity of the human eye to blurring is less for objects moving in a complex fashion, and the same is also true for a background [9]; and

b) *television pictures of moving objects are blurred partly* because of integration effects in the TV camera.

The subsampling pattern is a checkerboard and is frame interlaced. At the decoder, for stationary point, signals from the preceding frame are used as regenerated signals for unsampled points, and for moving points, spatially interpolated signals are used. To detect stationary points, we use the prediction errors of neighboring points (these points are sampled).

2) Mixed Mode: Normal sampling and subsampling are switched blockwise. Subsampling is used only in flat regions. The switching signal must be sent (1 bit/block) as described above.

E. Quantization and Threshold Processing

Quantization is performed in two steps. First, if and only if the level of the frame difference signal is below a limit L_N, the frame difference signal is level compressed at a ratio of 1/2. The coder thus acts as a noise reducer. In our experiments, L_N was set at two over the coding threshold.

Then the difference signal is quantized. We use three quantizers as shown in Table I. When the 5-bit quantizer is used, field prediction is used instead of frame prediction.

The switching of the quantizers is fieldwise.

A detail of the quantizer block is shown in Fig. 7. The coding threshold is modulated pel by pel in each field. In flat regions, the coding threshold is decreased by one or two (unit: 1/256) from the normal threshold $T_{cm\,ax}$. $T_{cm\,ax}$ is controlled fieldwise, and $T_{cm\,ax}$ is lowered periodically (5-9 fields) to decrease the remaining "dirty window" effect.

F. ΔM

In the mixed sampling modes, ΔM is used in a subsidiary role for very fast motion. This is DPCM of three levels. The characteristics of a DPCM quantizer are shown in Fig. 8. The switching between ΔM and interframe coding is blockwise.

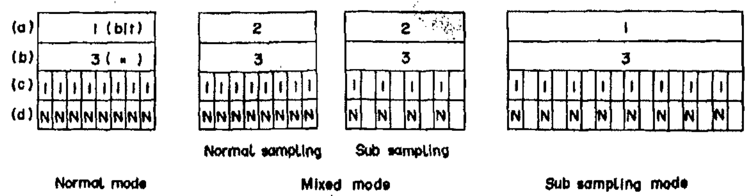


Fig. 6. Transmission formats. (a) Block flag. (b) Word length used. (c) Picture element flags. (d) Prediction error signal of N bits (N is the word length).

TABLE I
CHARACTERISTICS OF THE QUANTIZERS (THIS SHOWS POSITIVE SIDE ONLY). (a) 7 BIT QUANTIZER. (b) 6 BIT QUANTIZER. (c) 5 BIT QUANTIZER.

IN	OUT	IN	OUT	IN	OUT	IN	OUT
0	0	20~21	20	58~60	59	125~130	127
1	1	22~23	22	61~63	62	131~136	133
2	2	24~25	24	64~67	65	137~142	139
3	3	26~27	26	68~70	69	143~148	145
4	4	28~29	28	71~74	72	149~154	151
5	5	30~31	30	75~78	76	155~161	158
6	6	31~33	32	79~82	80	162~168	165
7	7	34~35	34	83~86	84	169~176	172
8	8	36~38	36	87~90	88	177~183	180
9	9	39~40	39	91~94	92	184~191	187
10	10	41~43	42	95~99	97	192~199	195
11	11	44~45	44	100~104	102	200~207	203
12~13	12	46~48	47	105~108	106	208~216	212
14~15	14	49~51	50	109~113	111	217~225	221
16~17	16	51~54	53	114~119	116	226~235	230
18~19	18	55~57	56	120~224	122	236~245	240
						246~255	250

(a)

IN	OUT	IN	OUT
0	0	36~40	38
1	1	41~46	43
2	2	47~52	49
3	3	53~59	56
4	4	60~66	63
5~6	5	67~75	71
7~8	7	76~84	80
9~10	9	85~94	89
11~12	11	95~106	100
13~14	13	107~118	112
15~17	16	119~132	125
18~20	19	133~148	140
21~23	22	149~165	156
24~27	25	166~184	173
28~31	29	185~205	192
32~35	33	206~228	213
		229~255	236

(b)

IN	OUT
0~1	0
2~4	3
5~9	7
10~13	11
14~18	16
19~25	22
26~33	29
34~41	37
42~51	46
52~61	56
62~73	67
74~86	79
87~100	92
101~114	106
115~130	121
131~142	137
142~255	154

(c)

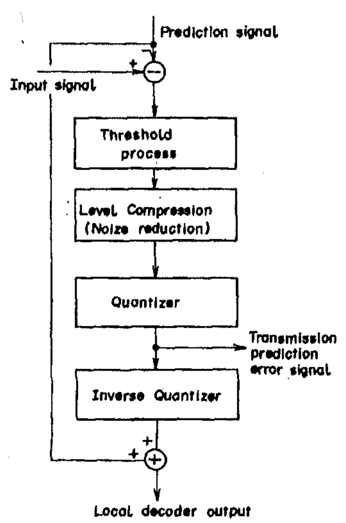


Fig. 7. Detail block of the quantizer.

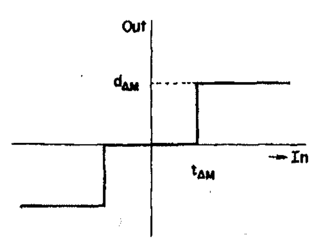


Fig. 8. Characteristics of ΔM quantizer. In our experiment, $t_{\Delta M} = 4, d_{\Delta M} = 3$.

The position for word-length signal (3 bits) is assigned to ΔM. The switching algorithm is as follows. When

$$\sum_{B_T} W_D(x_1, x_2) \geqslant T_{\Delta 1}$$

and

$$\max_{B_T} \{|A^N(x_1, x_2 - 1) - A^N(x_1, x_2)|\} \leqslant T_{\Delta 2}$$

Exhibit I Page 187

ΔM is used, where B_T is a transmission block, W_D is a level of frame-difference signal, and $T_{\Delta 1}$ and $T_{\Delta 2}$ are fixed thresholds. In our experiments, $T_{\Delta 1} = 16, T_{\Delta 2} = 5$.

G. Control of the Parameters

The coding threshold and sampling modes are switched to smooth the generation ratio of transmission information. The parameters to be controlled are threshold, quantizer, and sampling modes. It is difficult to control these parameters individually and so we use an order-table system. An example of such an order table is given in Table II. The order is selected fieldwise as a function of the occupancy of the buffer memory and the amount of generated information. The control algorithm is developed as follows. A number of selected order O is defined as a digitized function of the buffer memory occupancy B and the amount of information generated in the preceding field V^{-1} as

$$O = u \cdot F(B, V^{-1})$$

where u is a function that counts a fraction as a whole number and F is a combination of two functions, i.e.,

$$F = \begin{cases} F_1 & \text{for } O' > O^{-1} \\ F_2 & \text{for } O' \leqslant O^{-1} \end{cases}$$

where O^{-1} is the number of the preceding field and O' is the order number of the current field calculated using F_1. The function F_1 is

$$F_1 = \begin{cases} (B - b) \cdot O_M & \text{for } B - b > 0 \\ 0 & \text{for } B - b \leqslant 0 \end{cases}$$

where O_M is the number of the last order and b is a constant. We choose the value of b as 25 percent. F_2 is a function of B and V^{-1}, as

$$F_2 = \begin{cases} F_1 + \dfrac{(V^{-1} - R)}{2R} \cdot O_M & \text{for } V^{-1} - R > 0 \\ F_1 & \text{for } V^{-1} - R \leqslant 0 \end{cases}$$

where R is the amount of information transmitted in a field. Of course, the value of F_2 is limited by one.

The refresh interval is also controlled as a function of the occupancy of the buffer memory. The minimum interval is 1 s, and the interval is almost proportional to the occupancy of the buffer memory. In addition to this control format, if it exceeds 1 s and there is room in the buffer, the refresh is executed on a scene change.

IV. EXPERIMENTS

A. Experimental Setup

We built experimental "real-time" hardware as shown in Fig. 9. The experimental system corresponds to the inside of the broken line in Fig. 3. This experimental system only has a coder. Thus, we could not perform experiments on the effects of transmission errors. But all other experiments can be done using local decoder output signals. The experimental system encodes only half (ordinarily the right half) and on the display monitor, the coded/decoded half and the original half are displayed side by side.

TABLE II
ORDER TABLE

Sampling {N : normal, M : mixed, S : Sub}

Order Number	Quantizer	Coding threshold	Sampling
1	7 bit	1 LSD	N
2	7	2	N
3	7	3	N
4	6	3	N
5	6	4	N
6	7	2	M
7	7	3	M
8	6	3	M
9	6	4	M
10	5	4	M
11~12	5	5	M
13	5	6	M
14	6	3	S
15	5	6	S
16	5	7	S

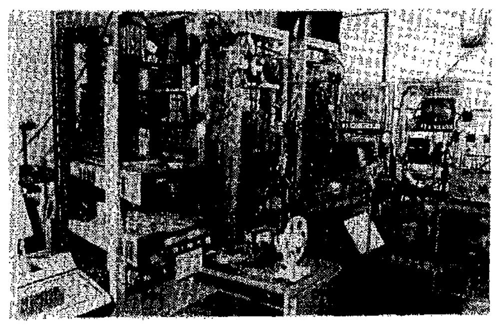


Fig. 9. View of the experimental system.

The original picture is coded 8 bits straight PCM, and the A/D converter has 10 bits resolution [10]. We think that such a configuration offers a very stringent test of the impairment produced by moving pictures.

The input video signals were monochrome decoded by a comb filter Y/C decoder with single one-line delay. The rejection ratio of the subcarrier of the decoder is about 40 dB.

The coding parameters are 1.6 bits/pel, 10.06 MHz sampling, and the buffer memory size corresponds to four frames.

B. Effects of Motion-Compensation

We show the effect of motion-compensation in Fig. 10. In this figure, the ratios of picture elements at which the prediction error signal exceeds a threshold t_c (active picture ele-

Exhibit I Page 188

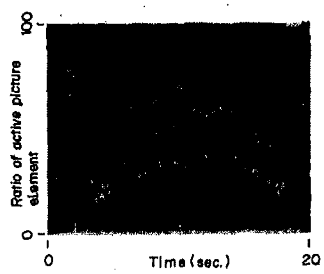


Fig. 10. Effects of motion compensation. Picture: Sport (Baseball). Upper: Uncompensated. Middle: Motion compensated (without iteration). Lower: Motion compensated (with iteration).

Fig. 11. An example of the "dirty window" effect. The right side of the picture is impaired. Coding threshold is 7.

ments) are shown. These data are derived from a 20 s scene showing a baseball pitcher. The upper curve corresponds to uncompensated prediction, the middle curve corresponds to noniterative (the system in which all $\hat{y}^{-1}$ in (8) are zero) motion compensation and the lower curve corresponds to the motion-compensated prediction.

C. Picture Impairment

Picture impairment is not severe under these conditions, but a slight degradation of picture quality can be observed. The degradations are as follows:

1) "dirty window" effect, although no block structure can be observed,

2) zig-zag of the vertical edge caused by the subsampling,

3) change of texture caused by the subsampling.

The "dirty window" effect is shown in Fig. 11. This is a very severe example. The threshold is fixed at seven, far from normal operating conditions. The other degradations may not be perceptible in printed pictures.

We observed many broadcast programs coded by the experimental system. The total length of the programs observed exceeded 200 h. These observations showed clearly that degradation of picture quality can be observed in the types of scene listed below. Other kinds of scene produced no problems:

1) scenes in which small objects are moving rapidly, for example, wide scene in which many children are jumping about,

2) scenes in which very small objects are moving at random, for example, a long shot of waves breaking on the seashore,

3) scenes in which the brightness level changes rapidly, for example, a stage decorated with flickering lights, and

4) scenes such as dissolves from a rapidly moving picture to a flat stationary picture, for example, a dissolve from a galloping scene to a picture of sky.

In ordinary broadcast programs, the sampling mode control will never enter the field repeat mode.

D. Operation Data

Difficult test scenes were selected by preobservation as described above. The duration of each scene was 20 s, and typical scenes are shown in Fig. 12. The S/N ratios were about 43 dB. Brief descriptions of the scenes follow.

Scene (a)–"Doll": Scenes of ordinary motion. The subjects are stuffed dolls, each moved by a man inside.

Scene (b)–"Sumo": Scenes of sumo, which is a kind of traditional Japanese wrestling. The picture includes rapid and violent motion.

Scene (c)–"Title": Title sequence for a drama. This includes fades as in 4) above and scenes like those in 2) above.

Scene (d)–"Stage": Staged scenes of a popular music number. This corresponds to 3) above.

Scene (e)–"Play": Scenes such as small children jumping about. This includes very high-speed zooming. This corresponds to 1) above. It also includes an artificial scene which flickers very rapidly.

Scene (f)–"Baseball": Scenes of a pitcher relieving. This includes extremely rapid motion.

In Fig. 13 we give data on the ratios of picture elements at which the simple frame difference signals exceed a threshold T (active picture elements) in order to show the degree of motion. In the figures, $T = 3$ and $T = 6$. The upper curves correspond to $T = 3$.

The time variation of the order numbers is shown in Fig. 14. These data show the effectiveness of the motion compensation. The order table used is as shown in Table II.

E. Subjective Assessment Test on Picture Quality

The picture quality of the scenes described above were evaluated in a subjective assessment test. The configuration of the test is as shown in Table III. We used a five-grade scale as shown in Table IV. The data were processed using the successive category rating method. The results are shown in Fig. 15. As may be seen from the figure, even for the most difficult picture (e), the degradation caused by the encoding was only about one grade.

By comparison, if uncompensated prediction is used, the picture quality is 2.3 and standard deviation is 0.95 for "Sumo." The experimental coder was adjusted to motion-compensated prediction, so it must be conceded that the setup was somewhat severe for conventional systems. However, even if this is taken into consideration, we can certainly achieve an improvement of almost two grades by using the motion-compensating system.

Exhibit I Page 189

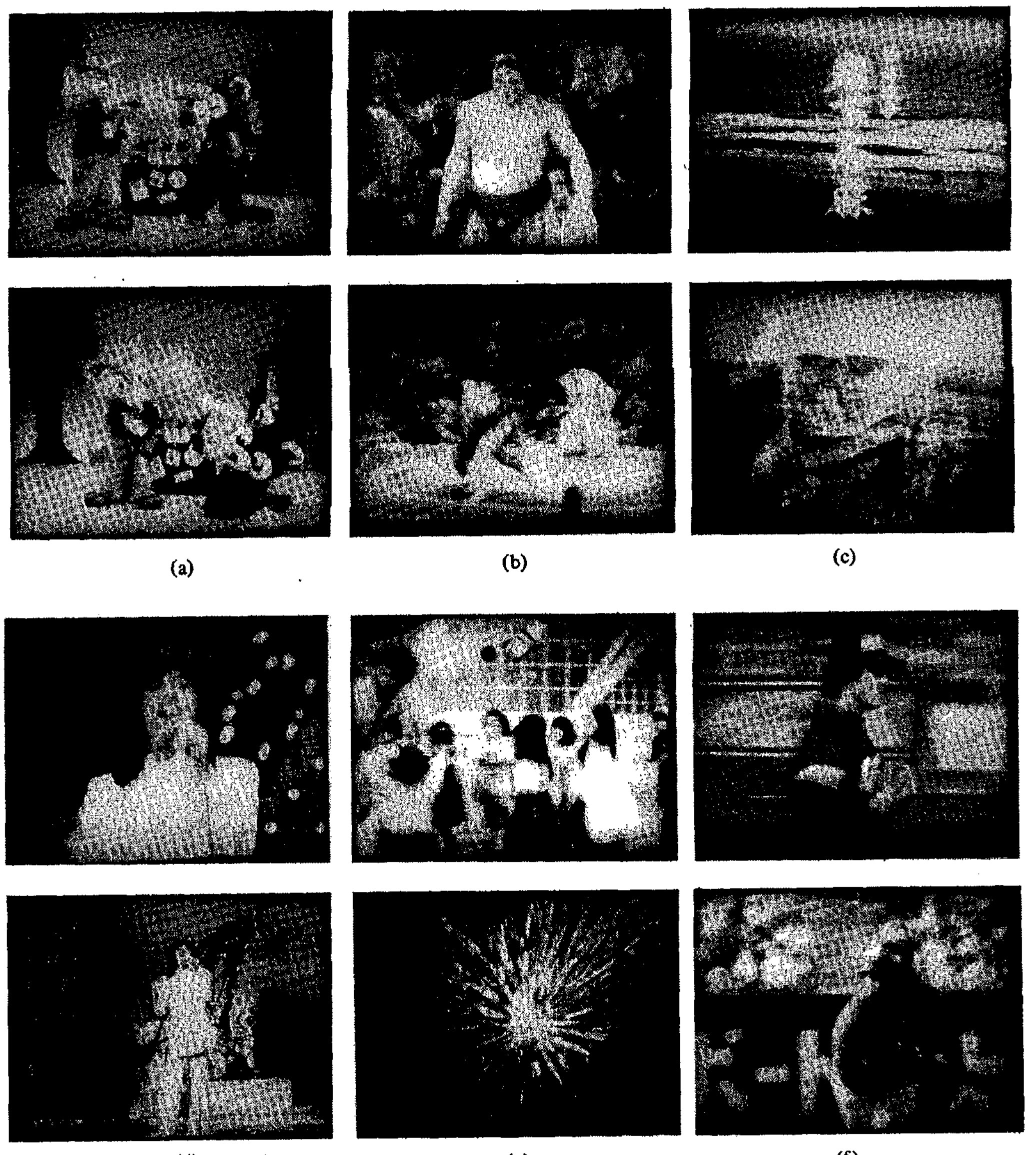


Fig. 12. Scenes used in the experiments. The duration is 20 s, time flows up to down. (a) Doll. (b) Sumo. (c) Title. (d) Stage. (e) Play. (f) Baseball.

Exhibit I Page 190

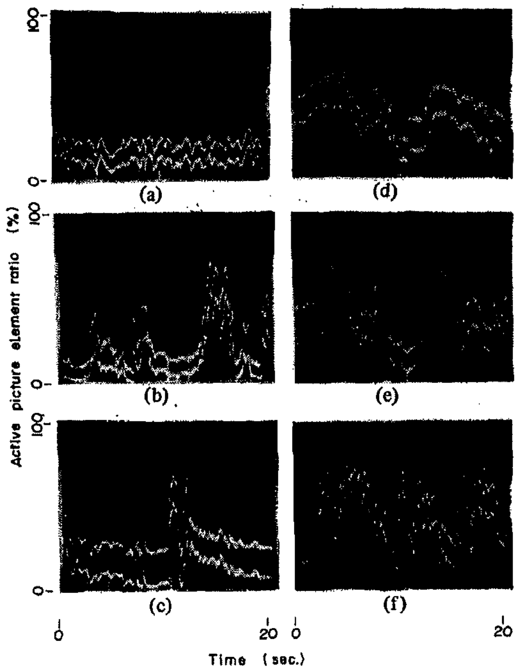


Fig. 13. Degree of motion of pictures. This shows the ratio of the active picture elements at which the simple frame difference signals exceed a threshold (T). The upper curves, $T = 3$; the lower curves, $T = 6$. (a) Doll. (b) Sumo. (c) Title. (d) Stage. (e) Play. (f) Baseball.

V. DISCUSSION

A. Transmission Bit Rate

We can estimate the transmission bit rate of the system as follows.

$$H_c = ((f_s \cdot 1.6 \cdot p) + c) \cdot (1 + E)$$

where p is the ratio of effective picture area, f_s is a sampling frequency, c the overhead required for the synchronizing signal in the PCM transmission, and E the overhead required for the error correction code. If $p = 0.8, f_s = 10.06$ MHz, $c = 0.3 \times 10^6$ (corresponds to 20 bits/line) and $E = 0.1$, we have $H_c = 14.5$ Mbits/s.

VI. CONCLUSION

We have established that a motion-compensated interframe coding system can be realized in practical apparatus. With such a system we can transmit at 1.6 bits/pel without an annoying impairment of the picture. With our experimental system, for the most difficult picture, picture quality was 3.7 on a five-grade impairment scale, and the degradation of picture quality was about one grade in comparison with the original picture.

By means of a simple extension of the monochrome system, we can apply the motion compensated interframe coding system to color television where a Y/C separate coding system is used. If, however, we use a composite coding system, we shall have to solve certain problems before we can handle color composite signals. The authors feel that the motion-

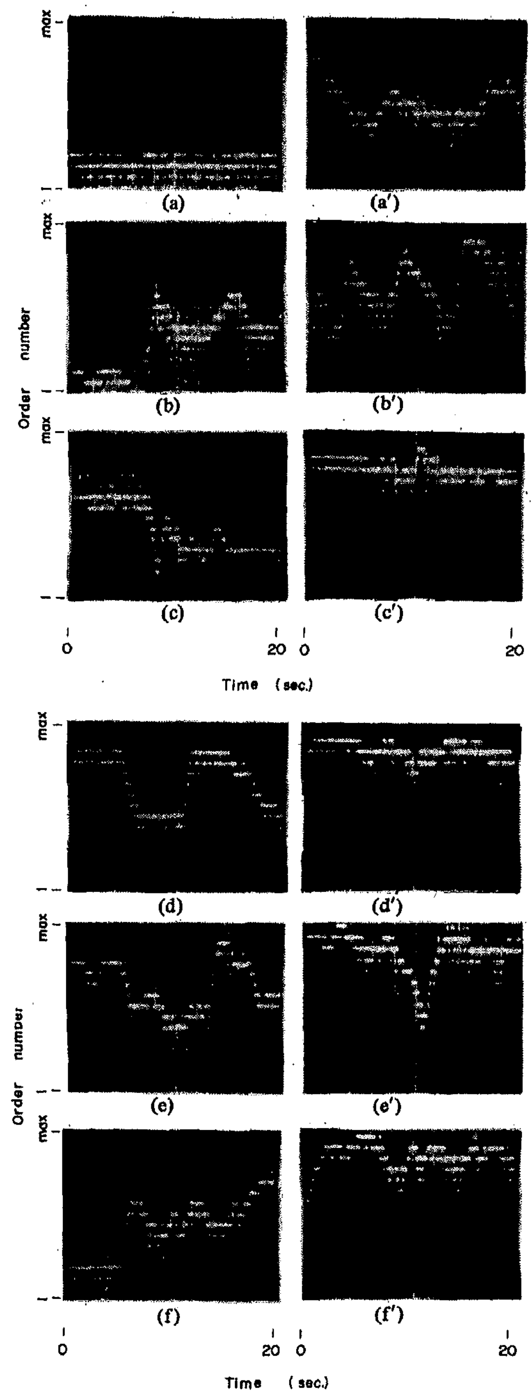


Fig. 14. Time variation of the order. (a) Doll. (b) Sumo. (c) Title. (d) Stage. (e) Play. (f) Baseball. The dashed line corresponds to the uncompensated prediction and the undashed line corresponds to motion compensated prediction. The maximum order number is 16.

TABLE III
CONDITIONS OF THE EXPERIMENT

Viewing distance	4 H
Maximum brightness	300 nit
Minimum brightness	3 nit
Ambient light	100 lux
Monitor size	18"
Phosphor	P_4
Subjects	52 Experts

TABLE IV
GRADES OF THE ASSESSMENT

For broadcasting, the picture impairments are······

5. not perceptible.
4. perceptible but not annoying.
3. slightly annoying.
2. annoying.
1. not usable because severely annoying.

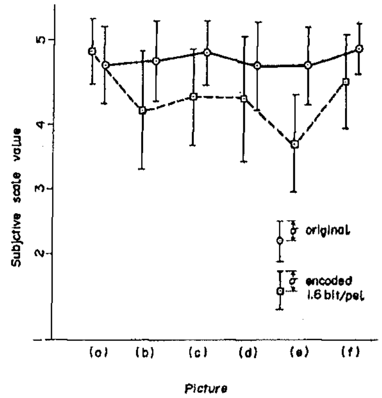


Fig. 15. Picture quality assessed in a subjective test. (a) Doll. (b) Sumo. (c) Title. (d) Stage. (e) Play. (f) Baseball.

compensation system should be Y/C separate type, even in a composite coding system.

APPENDIX

We denote the residual error of the determined motion vector as $\bar{r}$, where $\bar{r}$ is given by

$$\bar{r} = \hat{y} - \hat{y}_m$$

where $\hat{y}$ is the actual motion vector.

The power of the error in the motion-compensated frame-prediction is given by

$$\sigma_{\bar{r}}^2 = E[(A^N(x) - A_M{}^{N-2}(x))^2]$$

where $E[\cdot]$ denotes expectation and $A_M{}^{N-2}$ is the level of motion-compensated prediction signal. It is assumed that motion is uniform in the block. Thus, the level of the preceding frame is equal to the level of the shifted current frame. Then we have

$$\begin{aligned}\sigma_{\bar{r}}^2 &= E[(A^N(x) - A^N(x - \bar{r}))^2] \\ &= 2K_c{}^2\{1 - \psi(\bar{r})\} \qquad \text{(A1)}\end{aligned}$$

where ψ is an autocorrelation function and $K_0{}^2$ is the mean power of the picture signal. If we assume $K_0{}^2 = 1$, we have

$$\sigma_{\bar{r}}^2 = 2\{1 - \psi(\bar{r})\}.$$

If the level distribution of the differences between pairs of pixels of the picture is of exponential type and isotropic, we can assume that a probability density function of the frame level differences when the residual error of the detected motion is $\bar{r}$ as

$$P_{\bar{r}}(s) = \frac{1}{\sqrt{2}\sigma_{\bar{r}}} \exp\left(-\frac{\sqrt{2}}{\sigma_{\bar{r}}}|s|\right) \qquad \text{(A2)}$$

where s is a signal level of the frame difference.

The earlier assumption that we used is reasonable for a block in which the motion is simple, and the assumption that the autocorrelation function is of exponential type is also reasonable [7]. Using these assumptions and (A1) and (A2), we can deduce that D has a unique minimum, whether the measuring function is f_1 or f_2.

ACKNOWLEDGMENT

The authors wish to thank K. Hayashi, former Head of the Advanced Television Research Group, Dr. T. Fujio, Head of the Division, J. Ishida, the Senior Research Engineer, and the many researchers who took part in this work for their guidance, advice, and cooperation during the research.

REFERENCES

[1] F. Giorda and A. Roccii, "Bandwidth reduction of video signals via shift vectors transmission," *IEEE Trans. Commun.*, vol. COM-23, pp. 1002-1004, Sept. 1975.
[2] J. O. Limb and J. A. Murphy, "Estimating the velocity of moving images in television signals," *Comput. Graphics and Image Processing*, vol. 4, no. 2, pp. 311–327, 1975.
[3] Y. Ninomiya, "Motion correction for interframe coding systems," Inst. Electron. Commun. Eng., Tech. Rep. IE 78-6, 1978 (in Japanese).
[4] S. Brofferio and F. Rocca, "Interframe redundancy reduction of video signals generated by translated object," *IEEE Trans. Commun.*, vol. COM-25, pp. 448–454, Sept. 1977.
[5] A. N. Netravali *et al.*, "Motion compensated television coding—Part I," *Bell Syst. Tech. J.*, vol. 58, no. 3, pp. 631–670, 1979.
[6] E. R. Kreztmer, "Statistics of television signals," *Bell. Syst. Tech. J.*, vol. 31, no. 4, pp. 751–763, 1952.
[7] J. B. O'Neal, Jr., "Picture quantizing systems for the transmission of television signals," *Bell Syst. Tech. J.*, vol. 45, no. 5, pp. 689–721, 1966.
[8] Y. Ninomiya and Y. Ohtsuka, "A improvement of picture quality by 2nd order temporal prediction of interframe coding," Inst. Electron. Commun. Eng., Tech. Rep. IE79-87, 1979 (in Japanese).
[9] Y. Nimomiya and B. Prasada, "Some qualitative results of

Exhibit I Page 192

temporal masking in television pictures," in *Proc. 1979 Picture Coding Symp.*, Ipswich, England, 1979, paper 2.5.
[10] Y. Ninomiya, "An A/D converter with 10 bits/sample by 20 MHz sampling rate," *IEEE Trans. Commun.*, vol. COM-28, pp. 1–6, Jan. 1980.

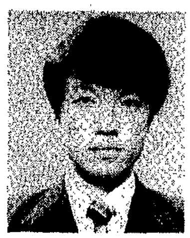
★

Yuichi Ninomiya was born in Tokyo, Japan, on September 17, 1942. He received the B.S. and Ph.D. degrees in electronic engineering from Keio University, Tokyo, Japan, in 1965 and 1973, respectively.

Since 1967 he has been engaged in research on laser applications and image processing and codings of television signals at the Technical Research Laboratories of the Nippon Hoso Kyokai (Japan Broadcasting Corporation), Tokyo. His research interests have included light modulation, deflection, and the materials to be used in laser applications. He is now conducting research on digital television systems.

Dr. Ninomiya is a member of the Institute of Electronics and Communication Engineers of Japan, the Institute of Television Engineers of Japan, and the Japan Society of Applied Physics.

★



Yoshimichi Ohtsuka was born in Aichi Prefecture, Japan, on December 17, 1948. He received the M.S. degree in electronic engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 1974.

Since 1977 he has been engaged in research on image processing and coding systems of television signals at the Technical Research Laboratories of the Japan Broadcasting Corporation, Tokyo. His research interests are image processing and picture coding systems. He is now conducting research on digital television systems.

Mr. Ohtsuka is a member of the Institute of Electronics and Communication Engineers of Japan and the Institute of Television Engineers of Japan.

Exhibit I Page 193