Novel Frame Interpolation Method for High Image Quality LCDs

Goh Itoh and Nao Mishima
Multimedia Laboratories, Corporate Research & Development Center, Toshiba Corp.
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki 212-8582, Japan, go.ito@toshiba.co.jp

Abstract

We developed a novel frame interpolation method to interpolate a frame between two successive original frames. Using this method, we are able to apply a double-rate driving method instead of an impulse driving method where a black frame is inserted between two successive original frames. The double-rate driving method enables improvement to the motion blur of LCDs caused by the characteristics of human vision without reducing the luminosity of the whole screen. Here, we report on the initial model of our frame interpolation method.

1. Objectives and Background

It is well known that the slow response time of liquid crystals (LC) causes motion blur of images on LCDs. We developed a driving method called LAO (Level Adaptive Overdrive) that reduces the LC response time to less than a single frame period (16.7ms) [1]. On the other hand, Kurita et al. analyzed a hold-type display method in which the picture is sustained during the frame period and showed that this display method degrades motion pictures due to the characteristics of human vision. Although moving objects are displayed at the same position during a frame period in the hold-type display method, the human eye continues to move in order to trace them. A difference exists between the viewpoint and the moving object. Since shifted images that depend on the difference are displayed on the retina, hold-type motion blurring of images occurs (Figure 1) [2]. This motion blur can be reduced by an impulse driving method or a double-rate driving method (Figure 2). In the impulse driving method, we insert a black frame between two successive original frames. In order to determine which driving method we should apply, 34 observers evaluated the simulated images subjectively with the universal motion picture simulator [3]. A still picture was scrolled horizontally as a motion picture.

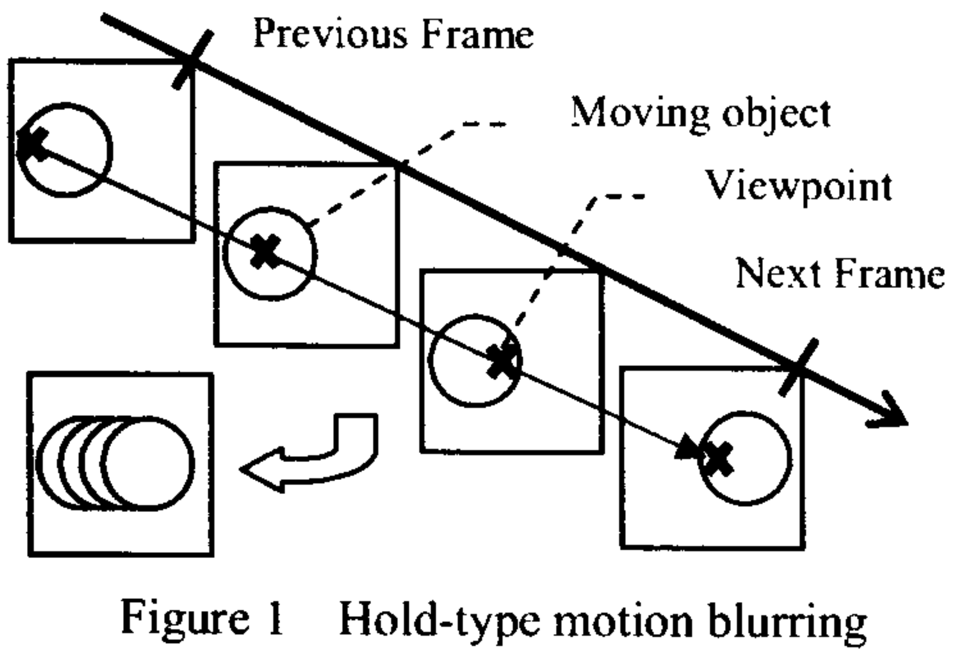


Figure 1 Hold-type motion blurring

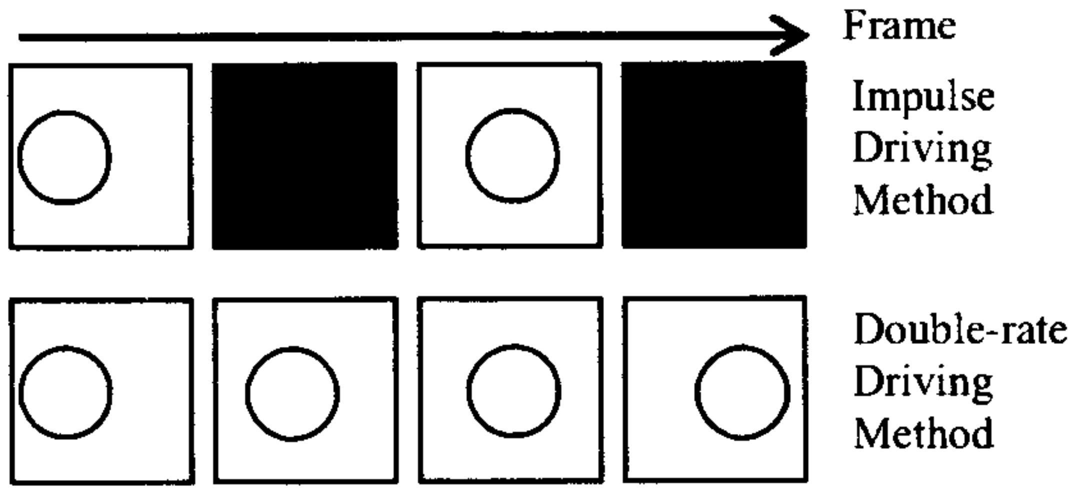


Figure 2 Driving method for blurring

The simulation conditions were as follows:
<CONDITIONS>
1) DISPLAY SPECIFICATIONS
Vertical frequency: 480 Hz
Horizontal frequency: 130 kHz
Image size: 7 inch diagonal
Image resolution: 320 x 240
2) HOLDING TIME (the number of sub-fields during sustaining the same picture): 1, 2, 4, 8 sub-fields
Defining the following equation as a holding rate, holding time of 1, 2, 4, and 8 sub-fields equals holding rate of 12.5, 25, 50, and 100 % respectively.
HOLDING RATIO (%H.R.) =
DISPLAYING PERIOD / (DISPLAYING PERIOD + NON-DISPLAYING PERIOD) * 100 --- (1)
3) LUMINOSITY
Maximum luminosity (white level) is the same for every display mode.
4) THE NUMBER OF FRAMES (fps)
For example, 240 fps means 240 picture frames per second.
5) OBJECTS MOVING SPEED: 16 degrees per second
We assumed that sport scenes have the objects moving speed of 10 to 20 degrees per second.
6) PICTURES: 2 kinds of CG picture & 2 kinds of natural picture

Figure 3 shows results of average subjective evaluation value for 4 kinds of picture. Subjective evaluation value depends on a motion blur and motion smoothness. Evaluation value "very good" means that a blur is almost equivalent to a still picture and objects move smoothly. For example, 240fps+50%H.R. and 120fps+25%H.R. are equivalent to 480fps+100%H.R. for motion picture image quality. In other words, they are the same in terms of hold-type motion blurring. Subjective evaluation value for 60fps+12.5%H.R. is, however, less than the above display mode, although a length of light emitting time for displaying the same frame is the same, 2.08ms. It indicates that the more frames there are per second, the higher the motion picture image quality is. We also found similar characteristics of subjective evaluation for other display modes such as a difference between 120fps+100%H.R. and 60fps+50%H.R.

Low frames per second leads to motion jerkiness caused by low correlation between a frame and the next frame. The motion jerkiness is likely to be remarkable for scrolling a clearer still image. Moreover, we confirmed that a motion blurring becomes large as holding ratio increases from 60fps+50%H.R. to 60fps+100%H.R.

Next, we investigated image quality with an actual LCD panel having high speed response property. The response time is about 16ms from white to black level. A black frame and a picture frame are displayed alternately at 60fps+50%H.R. Therefore, luminosity of a panel with 50%H.R. is half that with 100%H.R. Moreover, because LC response property is not impulse-type, luminosity of 60fps+50%H.R. is reduced to about 30% of 60fps+100%H.R. Figure 4 shows comprehensive subjective evaluation obtained by comparing 120fps+100%H.R. or 60fps+50%H.R. with 60fps+100%H.R. Subjective evaluation value of 60fps+50%H.R. was lower than that of 60fps+100%H.R. Judging from these results, we chose the double-rate driving method to realize 120fps+100%H.R.

2. Frame Interpolation Method

In the double-rate driving method, we insert a middle frame between contiguous frames. The middle frame is created by a frame interpolation method consisting of motion estimation and motion compensation. We constructed frame interpolation algorithms based on the block-matching algorithm used for MPEG. In the block-matching of MPEG, high correlation blocks are searched by motion estimation. There are various estimation functions [4], sum of absolute difference estimation function $E_{SAD}(d)$ is usually applied to MPEG, because the focus has been on reducing bit rates.

$$E_{SAD}(d) = \sum_{x \in B} \left| s(x-d,t) - s(s+d,t+1) \right| \quad \cdots (2)$$

$$d = \arg\min_{d \in W} E_{SAD}(d) \quad \cdots (3)$$

$x \in R^2$ represents the coordinate in frame $t \in R$ and $s(x,t)$ represents the intensity value at $(x,t)^T$. $d \in R^2$ represents a motion vector, B is a block and W is a search window.

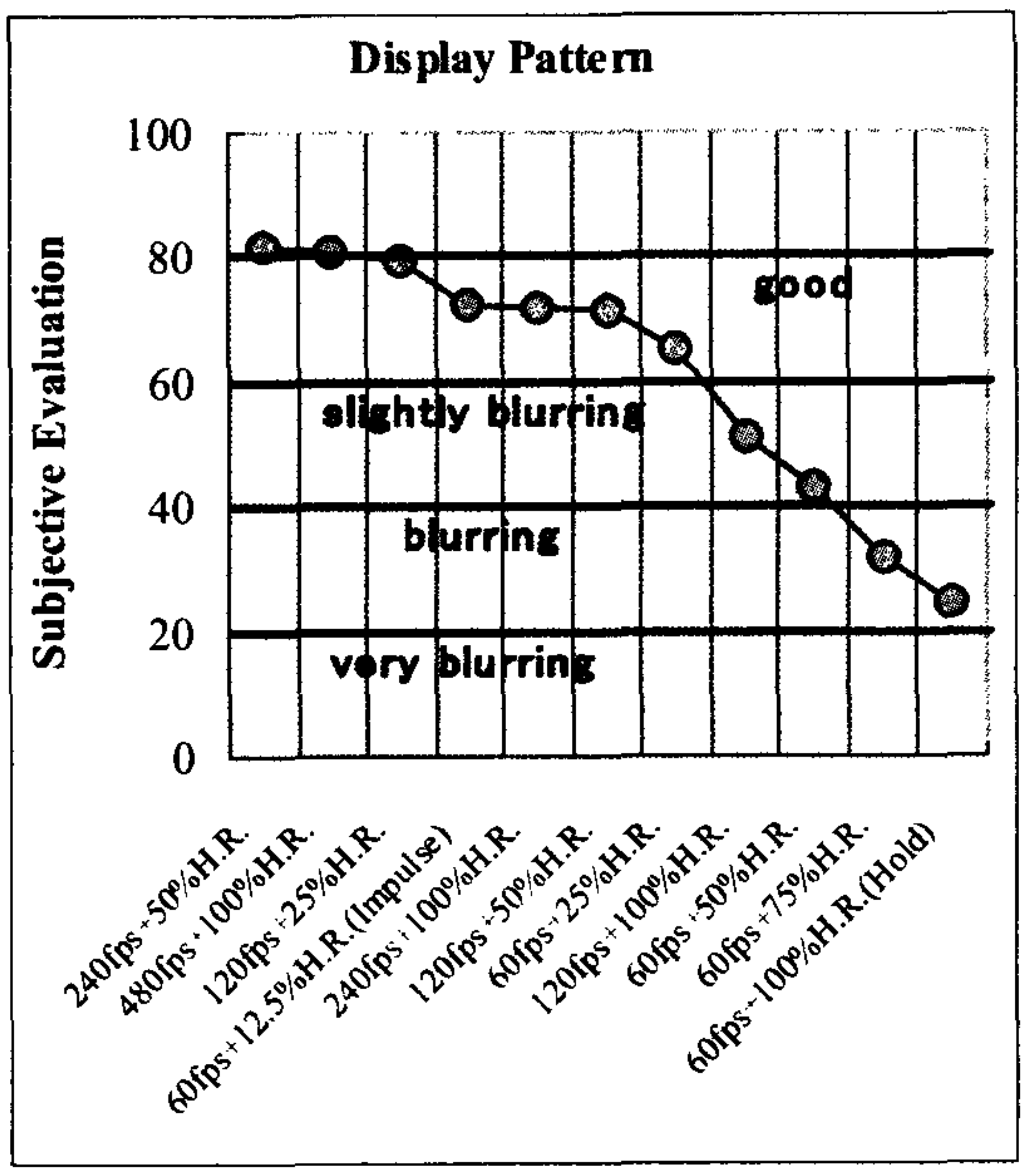


Figure 3 Subjective evaluation for motion picture quality with the simulation

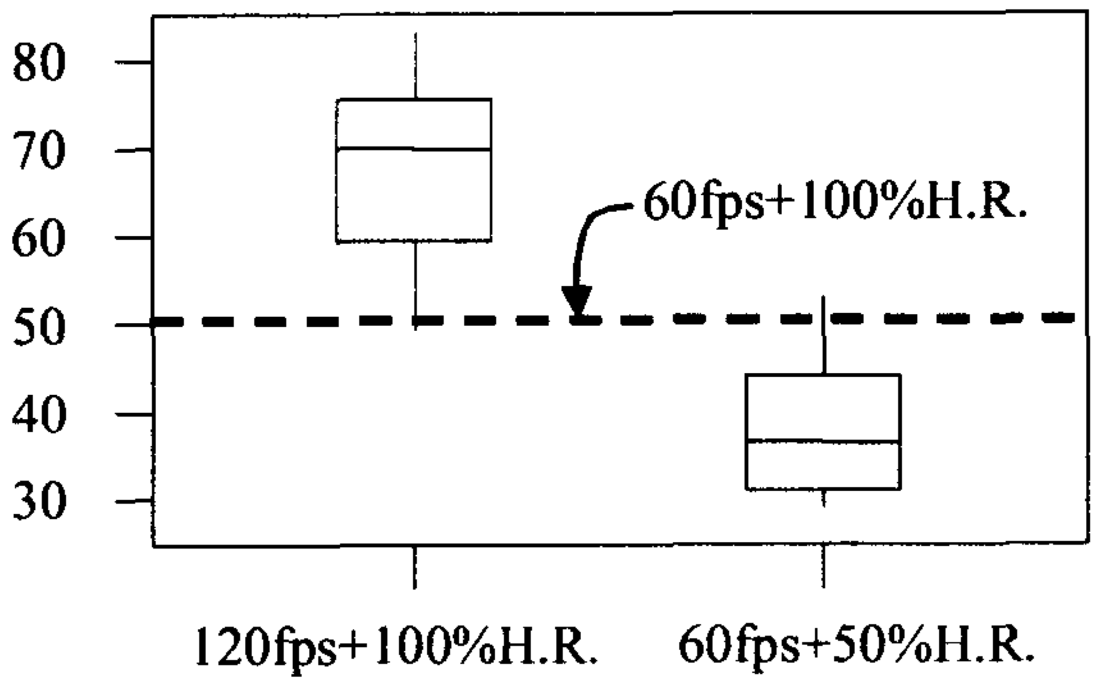


Figure 4 Comprehensive subjective evaluation for display mode with an actual LCD panel

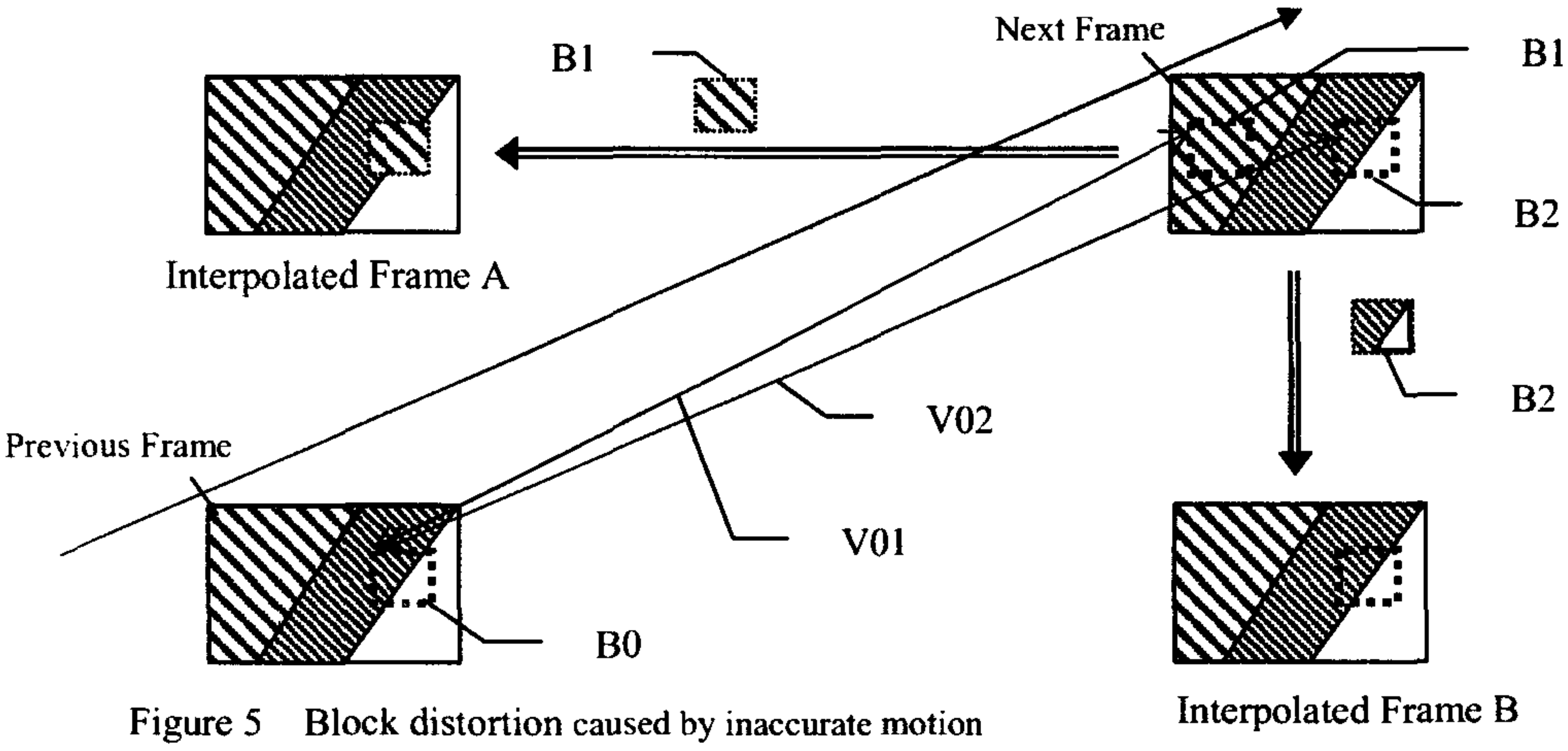


Figure 5 Block distortion caused by inaccurate motion compensation for boundaries of objects

Accurate motion estimation is very an important factor for motion compensation of the frame interpolation method. However, for example, different interpolated frames A and B with motion vectors V01 and V02, respectively, are able to be reconstructed as shown in Figure 5. There is block distortion caused by low correlation block B1 in the interpolated frame A. Therefore, we needed another estimation function to focus on image texture. Maximum matching pixel count estimation function $E_{MPC}(d)$ depicted in Equation (4) to (6) is as follows.

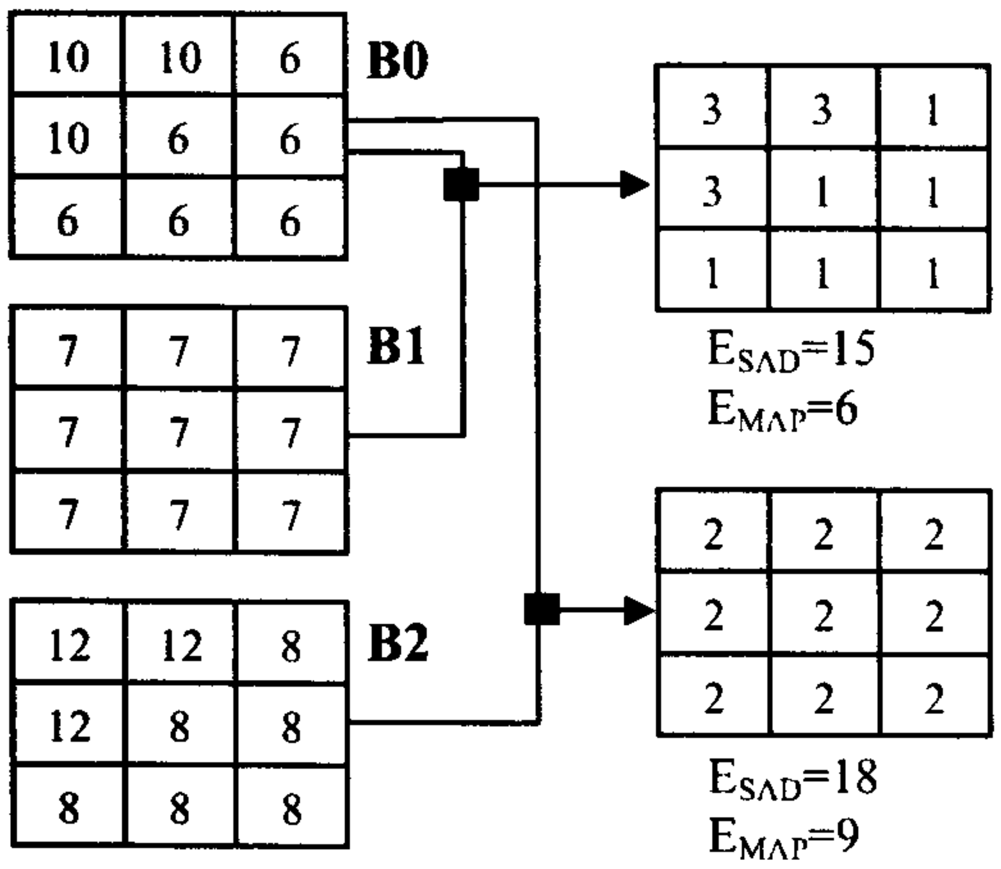
$$E_{MPC}(d) = \sum_{x \in B} \sigma(x,d) \quad \cdots(4)$$

$$\sigma(x,d) = \begin{cases} 1 & |s(x-d,t) - s(x+d,t+1)| \leq T1 \\ 0 & otherwise \end{cases} \quad \cdots(5)$$

$$d = \arg\max_{d \in W} E_{MPC}(d) \quad \cdots(6)$$

Figure 6 shows the example of block candidates B0, B1 and B2 in Figure 5, and results of $E_{SAD}(d)$ and $E_{MPC}(d)$ (Searching Threshold:*T1* in equation (5) = 2).

Although $E_{SAD}(d)$ between B0 and B1 is smaller than $E_{SAD}(d)$ between B0 and B2, B1 is a raster texture. On the other hand, although average intensity value of B2 became higher than that of B0, $E_{MPC}(d)$ between B0 and B2 is larger than $E_{MPC}(d)$ between B0 and B1. $E_{MPC}(d)$ is likely to be more dependent on image texture than $E_{SAD}(d)$.



B0

10	10	6
10	6	6
6	6	6

B1

7	7	7
7	7	7
7	7	7

B2

12	12	8
12	8	8
8	8	8

3	3	1
3	1	1
1	1	1

E_{SAD}=15
E_{MAP}=6

2	2	2
2	2	2
2	2	2

E_{SAD}=18
E_{MAP}=9

Figure 6 Correlation between block candidates obtained by calculating $E_{SAD}(d)$ or $E_{MAP}(d)$

In addition to the above matching characteristics, to avoid block distortion caused by inaccurate motion compensation for boundaries of moving objects when two or more moving objects exist in a block, we applied an area segmentation technique to the basic block-matching algorithms. As shown in Figure 7, the first block matching between two contiguous frames was performed geometrically and symmetrically centered on an interpolated frame. Secondly, we divided a block into a high correlation area (dotted area) and a low correlation area (black area) depending on a segmentation threshold *T2* which can be the same as value *T1* in equation (5). Thirdly, in the second block matching, high correlation blocks (hatched area) whose shape is the same as the low correlation area are searched between previous frame and next frame using the same searching algorithm as for the first block matching. Next, the first motion vector detected by the first block matching was assigned to the high correlation area and the second motion vector detected by the second block matching was assigned to the low correlation area. Then, image data of pixels that motion vectors are assigned to was pasted to an area in the interpolated frame according to each motion vector. Finally, an area in the interpolated frame that image data was not pasted to was allotted from peripheral pixels or contiguous frame. Figure 8 shows pixel interpolation for the area without image data obtained by first and second block matching. Gray pixel in the interpolated frame is calculated by non-linear averaging such as median filter or linear averaging. Image data, using non-linear averaging, are retrieved from peripheral pixels (hatched pixels) with image data obtained by first and second block matching and contiguous pixels (dotted pixels) which are pointed to by first and second motion vectors on original frames. If there is no image data in a peripheral pixel, we ignore the pixel. There are several pixel interpolation methods. We intend to investigate the other method based on motion objects segmentation.

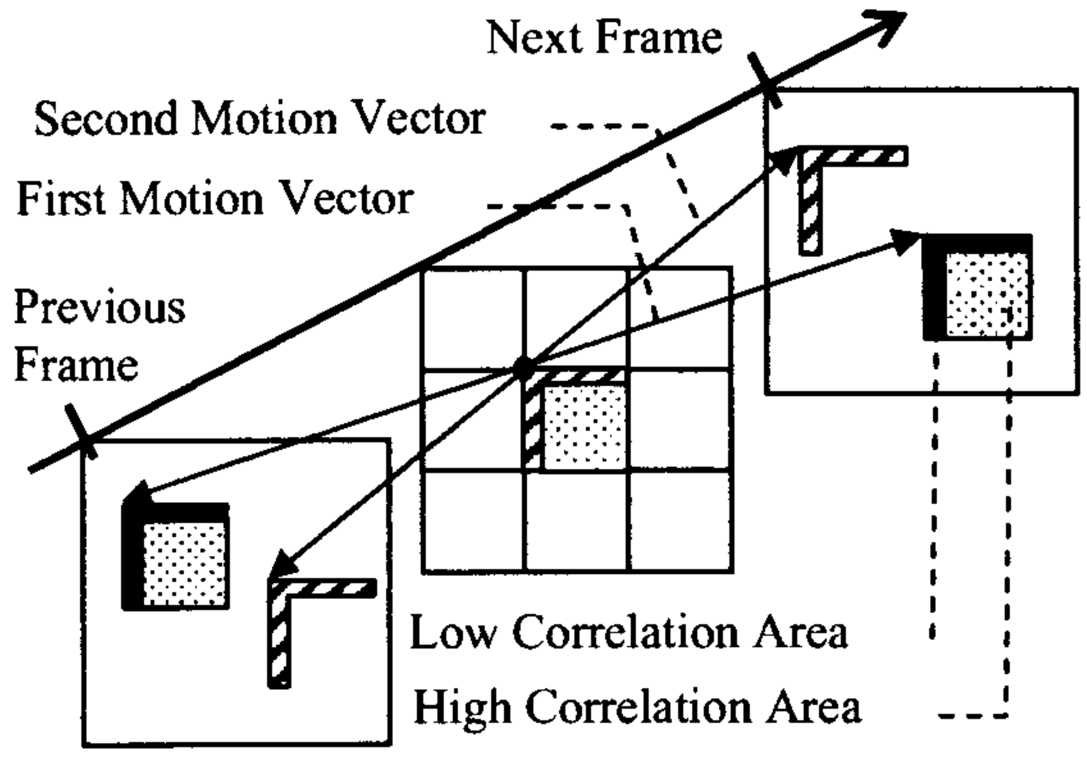


Figure 7 Proposed interpolation method

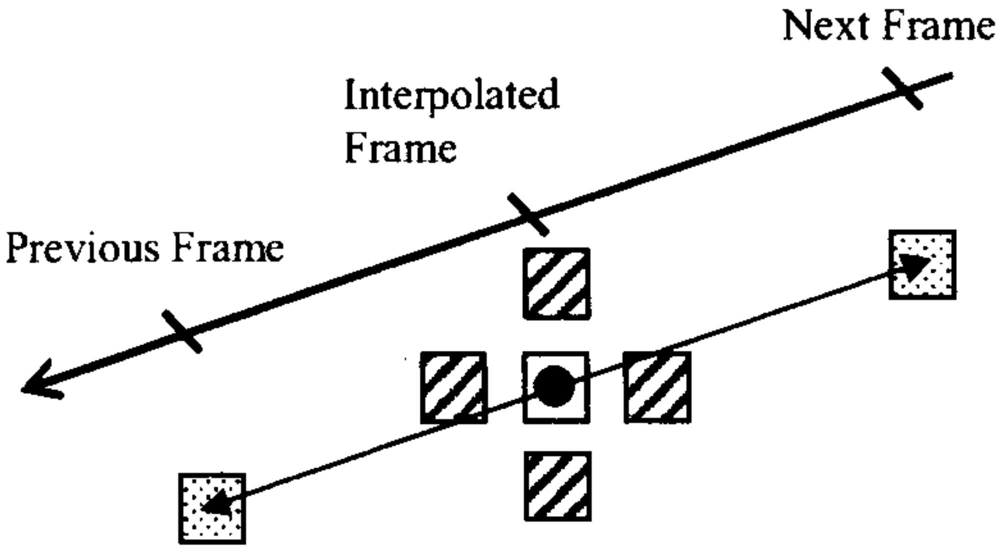
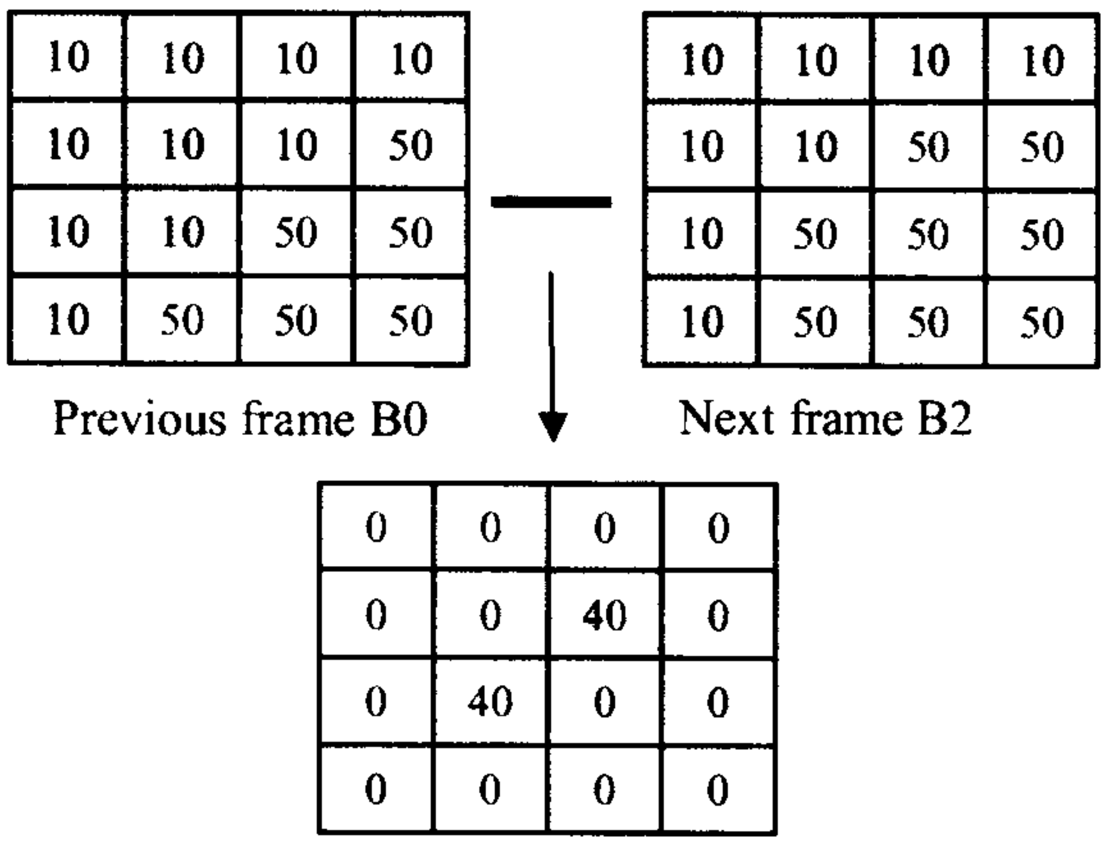


Figure 8 Pixel interpolation

Here, we examine the thresholds *T1* and *T2*. *T1* is the searching threshold for motion estimation and *T2* is the segmentation threshold for motion compensation. It is not necessary for *T1* to be equal to *T2*. If *T1* is large, block matching does not depend on texture and if *T1* is small, block matching is sensitive to slight intensity change such as noise. On the other hand, if *T2* is large, motion compensation is insensitive to texture and noise and if *T2* is small, the number of pixels without image data becomes large. Too small *T1* and too small *T2* increase pixels without image data, which leads to salt and pepper noise due to low correlation between a pixel with interpolated image data and a peripheral pixel with obtained image data. For example, if the object deforms slightly, the edge of the object shifts around a few pixels. B0 in the previous frame and B2 in the next frame are retrieved and then the absolute difference value per pixel is calculated, two pixels (dotted pixels) in the center of a block become pixels without image data as shown in Figure 9. These pixels usually appear on the edge of objects, because the luminosity difference between neighboring pixels in the edge region is generally higher than that in the raster region. Therefore, we applied a different segmentation threshold between the edge region and the raster region. The threshold *T2* for the edge region is higher than that for the raster region.



10	10	10	10
10	10	10	50
10	10	50	50
10	50	50	50

Previous frame B0

10	10	10	10
10	10	50	50
10	50	50	50
10	50	50	50

Next frame B2

0	0	0	0
0	0	40	0
0	40	0	0
0	0	0	0

Absolute difference between pixels

Figure 9 Salt and pepper noise

3. Results

Figure 10 shows an interpolated picture realized by using our interpolation method. Salt and pepper noise remained between characters and a portion of a character such as "Test" on the right side disappears in the "Bronze" picture. Similarly, a foot of an umpire or a player disappears in the "Hockey" picture. Moreover, "COFFEE" on the left side has been broken. These image degradations mainly result from inaccuracy of motion estimation. We have used image quality measures: Subjective Mean Squared Error (SMSE) [5]. The SMSE which we used is defined by:

$$SMSE = \left(\frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} |g(x,y)|^{3} \right)^{1/3} \quad \cdots (7)$$

$$g(x,y) = M_{2,2}(s_{ref}(x,y)) - M_{2,2}(s_{test}(x,y)) \quad \cdots (8)$$

$M_{2x2}(s(x,y))$ denotes a linear average in a 2x2 pixel window to describe the low-pass filter of the eye. Image quality has been improved by adding options to the interpolation method as shown in Figure 11. Here, we set image data of pixels to which no motion vector is allotted to 0 in Without Pixel Interpolation. However, more improvement is necessary for reducing SMSE to less than 10.

4. Conclusion

In this paper we demonstrated an initial model of the frame interpolation method between two successive original frames and we verified the validity of our method. The more improved and robust method leads to realization of the double-rate driving method for improving motion blur for LCDs.

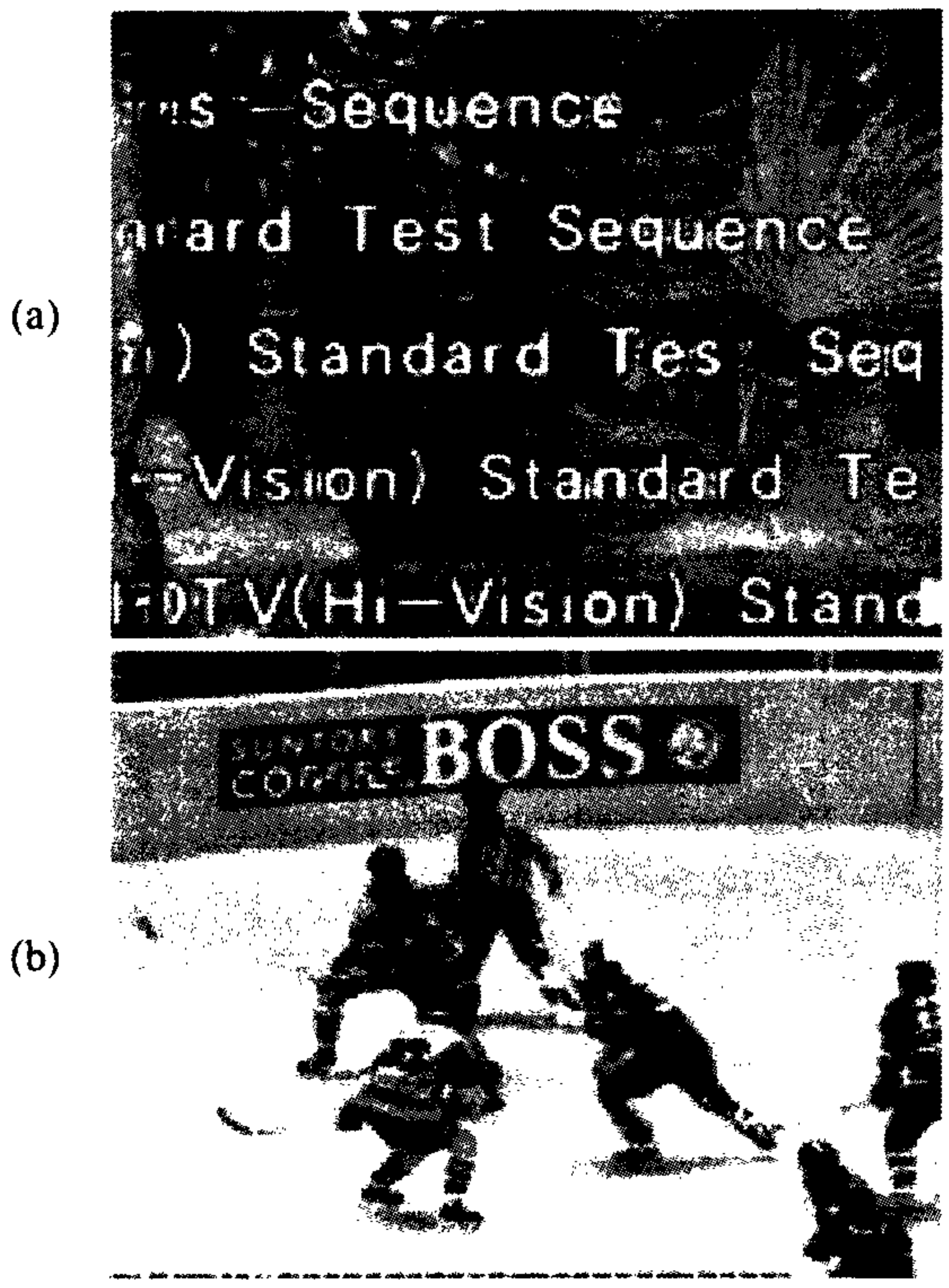


Figure 10 Interpolated pictures: (a) Bronze; characters move right to left on a background image with natural plants. (b) Hockey; Players move in various directions.

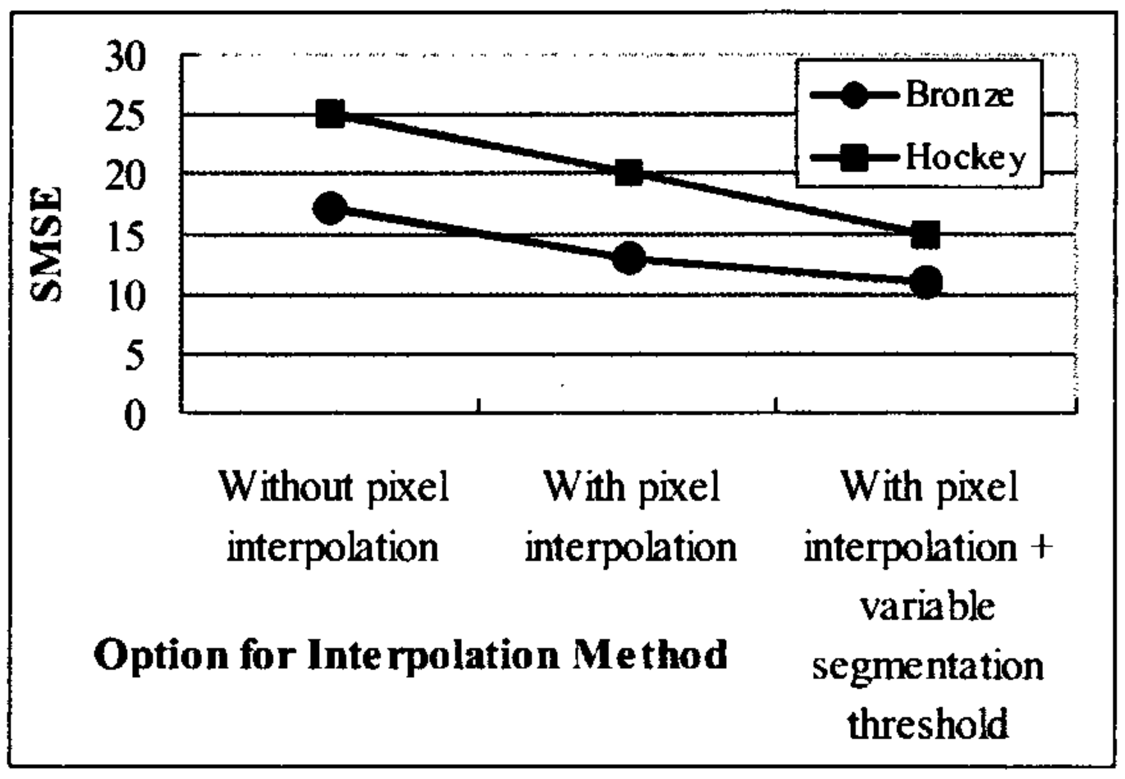


Figure 11 SMSE of interpolated pictures

5. References

[1]H. Okumura and H. Fujiwara, SID Vol.1 No.3, P335-339 (1993)

[2]H. Ishiguro, T. Kurita, Tech. Rep. of IEICE, EID96-4, P19-26 (1996)

[3]G. Itoh, M. Baba and H. Okumura, Asia Display/IDW'01, P1633-1636 (2001)

[4]A. M. Tekalp, "digital video processing", Prentice Hall, Chapter 6 (1995)

[5]H. Marmolin, IEEE Trans. on System, Man and Cybernetics, Vol.16/3, P486-489 (1986)