Table 6. Spatial Blurring Features For Figure 13

Scene	M-PSDI	SD-PSDI	RMS-PSDI	NPGT-PSDI
Top Row (DS1)	17.5	35.2	39.4	5443
Second Row (1/2 DS1)	19.0	37.9	42.4	6562
Bottom Row (1/4 DS1)	22.5	47.4	52.5	9544

Table 7. False Edge Features For Figure 13

Scene	M-NSDI	SD-NSDI	RMS-NSDI	NPLT-NSDI
Top Row (DS1)	-9.0	20.7	22.6	1138
Second Row (1/2 DS1)	-9.6	22.5	24.4	1577
Bottom Row (1/4 DS1)	-11.5	25.0	27.5	1989

2.7 Jerkiness Feature Using Position Errors

Jerkiness is a video teleconferencing/telephony artifact in which the original smooth and continuous imagery motion is perceived as a series of distinct snapshots at the output (see Table 1). Jerkiness is normally present when a codec data compression algorithm achieves data compression by elimination of fields or frames. The number of fields and/or frames that are eliminated (not transmitted) is not necessarily guaranteed to be an accurate measure of jerkiness. Sophisticated coding algorithms can update different portions of the image at different frame rates and even interpolate missing frames to achieve smooth motion effects. Jerkiness is present when the position of a moving object within the video scene is not updated rapidly enough. Section 2.7.1 proposes a measure for jerkiness based on injecting a video scene containing a moving object, and then measuring the object's position

errors in the output video. The technique is general enough to use an arbitrary object which is undergoing translational motion. A stored image of the stationary object is required to implement the technique. Although the jerkiness feature presented in section 2.7.1 is very accurate, there is one shortcoming. The feature cannot, in general, be extracted from arbitrary video scenes. Section 2.8.1 of this report will propose another measure of jerkiness which can be extracted from any video scene.

For moving objects, the proposed measure of jerkiness complements the previously proposed measures of spatial blurring in that the jerkiness feature measures temporal positioning accuracy of the object while the spatial blurring features measure the spatial resolution of the object. Data compression of motion video often involves a tradeoff between allocating bits to the temporal or spatial attributes of moving objects. The ability to measure separately the temporal and spatial attributes raises the possibility of tailoring performance specifications to the application. For example, consider the application of VTC/VT for trouble shooting circuit diagrams. In this application, high spatial resolution of the circuit diagram (assumed to be mostly stationary) is much more important than having a moving pointer (such as a finger or pen) seen as smooth and continuous. In other applications involving head and shoulders video teleconferencing, having less jerkiness may be more important than having high spatial resolution.

2.7.1 Feature Extraction Technique

A feature for estimating the jerkiness in sampled video imagery can be obtained using very simple image processing techniques. The jerkiness feature can be extracted by injecting a video scene that contains a moving object. The horizontal and vertical motion of the object is then tracked for the output imagery. Comparing the vertical and horizontal motion trajectories of the output to the input, a useful measure of jerkiness is obtained. The input motion trajectory can be obtained from processing the input video scene, or from a priori knowledge (since the test signal is known). In this manner, the amount of jerkiness in the output imagery relative to the input imagery can be determined. Several

steps are required to apply the technique. Time alignment of the input and output video scenes before processing is not required. The alignment will be performed on the input and output motion paths (see step 5), rather than on the input and output video scenes. The following steps are applied to extract the feature.

1. Stationary reference object

   A stationary reference image of an object against a uniform background is stored. This reference image of the object will be used to track motion jerkiness. The technique is general so that any non-rotational, non-growing or shrinking object may be used. For simplicity, a black ball on a white background was used for the experiments presented later in this report.

2. Moving reference video scene

   Successive frames of the object in step 1 above are generated with the object moving (translating in vertical and horizontal positions). The object may be moved horizontally, vertically, or diagonally depending upon whether one desires to test the jerkiness in the horizontal, vertical, or diagonal directions. The object may also be moved at different velocities to test the jerkiness over a wide range of motion in the video scene. In this manner, a video scene is generated that contains an object moving according to some known motion path (the vertical and horizontal positions of the object are known for each video frame).

3. Output video scene

   The generated video scene from step 2 above is injected as the test signal. The output video scene is recorded or frame grabbed into the video quality assessment system. For greater accuracy, each field (1/60th of a second) was recorded for the experiments presented later in this report.

4. Output motion path

The vertical and horizontal positions of the moving object are obtained by correlating (see Oppenheim and Schafer, 1975) each video frame of the output video scene (from step 3 above) with the reference object (from step 1 above). In this manner, the vertical and horizontal motion paths of the moving object are found for the entire output video scene. Correlation yields a very robust and accurate estimate of the moving object's position. However, correlation is also computationally expensive. A computationally more efficient, but less accurate, method of tracking the moving object's position is available if the object is against a black background. Then one could obtain the object's motion path by computing the centroid of the object for each video frame (Tzafestas, 1986). Nevertheless, the correlation method was used for the experiments presented later in this report.

5. Aligned output motion path

The output motion path of the object (from step 4) is aligned with the true motion path (from step 2). Alignment of the input and output motion paths is required to compensate for absolute video delay of the device under test. The alignment procedure used here corresponds to what a viewer would observe if that viewer were insensitive to the absolute video delay. The best alignment of the output motion path to the input motion path is simply that which produces the smallest average sum of the squared vertical and horizontal position errors (the sum of the squared position errors is first performed over all frames of the video scene, then this sum is divided by the number of frames in the video scene). The jerkiness feature is then calculated as the square root of this average sum of the squared position errors. A mathematical definition for this jerkiness feature, henceforth called temporal root mean square position error (TRMS-PE), is given in equation 15 of Appendix A.

2.7.2 Sample VTC/VT Results

The notion of testing the jerkiness of motion video first occurred when the output of a VTC/VT codec was monitored at bit rates on the order of DS1. An object that moved across the field of view of the camera did not seem to move as smoothly after the scene had passed through the codec. A quantification of how jerky the distortion mechanism was and how it varied with code rate and speed of the object was sought.

An ideal test signal for jerkiness would be a computer generated scene of an object moving at a constant speed across the screen at a specified angle (horizontally, vertically, diagonally). Due to equipment limitations, test scenes were generated using a black ball suspended by a long pendulum (about 15 feet) against a backlit (white) background. Since only a small portion of the center part of the swing was used, the ball's speed and angle were approximately constant.

A black ring was placed on the backlit background so that background movement due to imperfections in the test setup or recorders could be detected and taken into account. For very stable recorders or computer generated scenes, the black ring would not be necessary.

To generate test scenes of different speeds, the ball was dropped from different heights. To generate test scenes at different angles, the camera was tilted to the appropriate angle. In this manner, test scenes were generated for horizontal and 45 degree angles at several different velocities (ball heights). Three consecutive swings from each ball height were captured into the computer. For each scene, every set of two fields that could be displayed on the video cassette recorder in still frame mode was captured and stored in a file for later processing. Images were grabbed for every NTSC field increment of the recorder (1/60 second). Although the speed of the consecutive swings for each ball height was slightly decreasing, the motive was to establish the accuracy and repeatability of the jerkiness measurement by examining three independent trials at each ball speed. The following scenes with horizontal motion were captured into the computer and analyzed:

1. Nine original reference scenes (three consecutive swings of the ball for each of three different ball heights or speeds).

2. Nine degraded codec output scenes at the ball's fastest speed (DS1, 1/2 DS1, and 1/4 DS1 code rates for the three consecutive swings at the fastest speed).

3. Six degraded codec output scenes at the ball's medium speed (DS1, and 1/4 DS1 code rates for the three consecutive swings at the medium speed).

4. Six degraded codec output scenes at the ball's slowest speed (DS1, and 1/4 DS1 code rates for the three consecutive swings at the slowest speed).

The following scenes with 45 degree diagonal motion were captured and analyzed:

1. Three original reference scenes (three consecutive swings of the ball at the fastest speed).

2. Nine degraded codec output scenes at the ball's fastest speed (DS1, 1/2 DS1, and 1/4 DS1 code rates for the three consecutive swings of the ball at the fastest speed).

The horizontal and vertical motion paths of the ball for each scene listed above were obtained by correlating a stored reference ball with each image of the video scene. Possible movement of the background (which contained a black ring) due to imperfection in the test setup was detected by correlating a stored reference ring with each image of the video scene. The motion of the background in the test setup was found to be on the order of one or two pixels and hence was neglected.

Figure 14 shows four sequential images grabbed (every 1/60th of a second from left to right) at the various bit rates for a horizontally moving ball. The top row in Figure 14 shows four consecutive field increments of the original NTSC signal, the next three rows show the corresponding codec outputs at bit rates of DS1, 1/2 DS1, and 1/4 DS1, respectively. For viewing convenience, single-frame temporal alignment has been applied to the video in Figure 14. Note that the second and third images in each row of the codec output are correctly aligned with

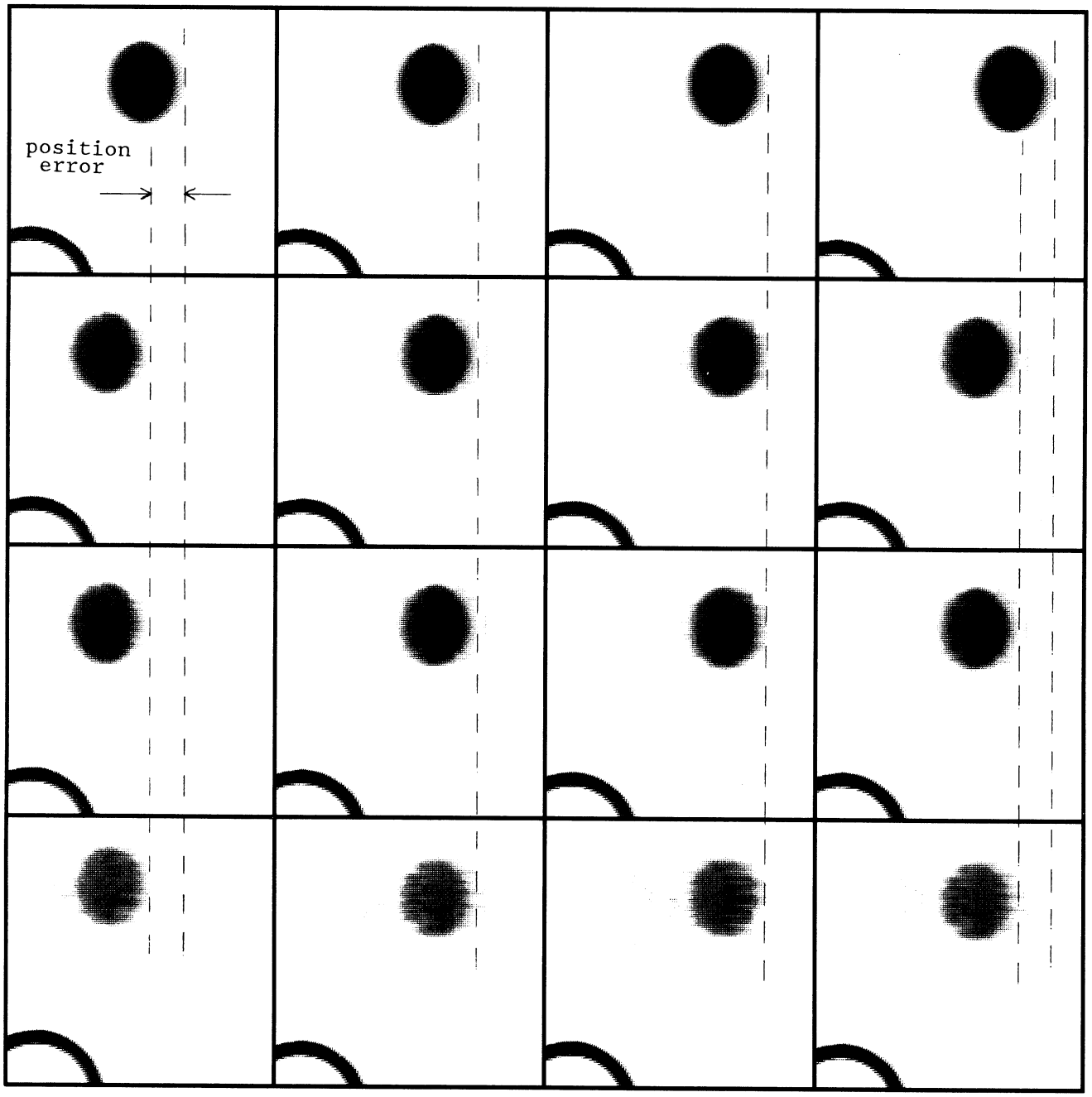

Figure 14. Four sequential VTC/VT images for a horizontally moving ball. Images were grabbed every 1/60 second from left to right. Original (first row), DS1 (second row), 1/2 DS1 (third row), 1/4 DS1 (last row).

the NTSC input. For each bit rate, the ball is identically positioned, but this positioning is not the same as the input in the first and fourth codec output images. In addition, for each bit rate, the ball in the fourth codec output image appears to have backed up while the original continues to advance from left to right. The reason for the strange positioning of the moving ball in the codec output video will be explained below. Figure 15 shows a portion of the diagonal test data with the ball moving from the upper left to the lower right. The format of Figure 15 is the same as that of Figure 14.

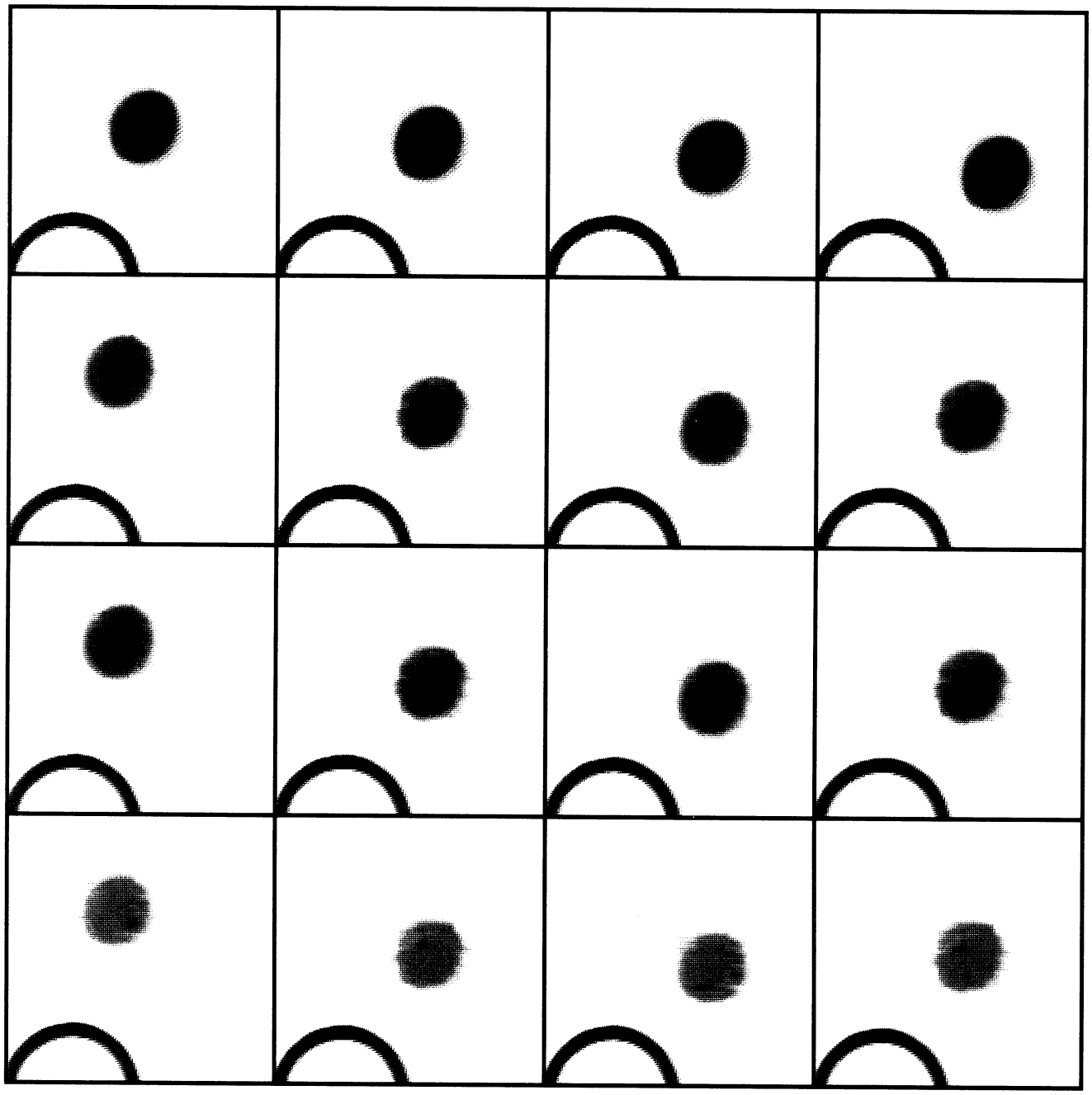

Figure 15. Four sequential VTC/VT images for a diagonally moving ball. Images were grabbed every 1/60 second from left to right. Original (first row), DS1 (second row), 1/2 DS1 (third row), and 1/4 DS1 (last row).

Figure 16 shows the horizontal and vertical positions of the ball as a function of field number for fast motion at the horizontal angle. The ball positions are plotted for the original NTSC scene and for a code rate of DS1. Figure 17 shows the ball positions for fast motion at the diagonal angle of approximately 45 degrees. In Figures 16 and 17, the codec very accurately positioned the ball for two consecutive fields, but then, to save on transmission, simply repeated these two fields before accurately placing the ball again. This omission and repetition of every other frame caused the backup mentioned earlier in Figures 14 and 15. Examining Figure 14, the ball position in the first DS1 output image corresponds to field number 3 in Figure 16. In the second DS1 output image, the ball jumps a large distance to field position 4 in Figure 16. The ball in the second and third DS1 output images was accurately placed (corresponding to field numbers 4 and 5 in Figure 16). Then, the ball in the fourth DS1 output image (field number 6 in Figure 16) backed up because the codec output the same field that occurred earlier in time (field number 4 in Figure 16). Thus, the fourth DS1 output image in Figure 14 was identical to the second DS1 output image (since field number 6 is identical to field number 4 in Figure 16).

In order to measure the TRMS-PE feature, the input and output motion paths had to be aligned according to processing step 5 of section 2.7.1. Figure 18 shows the aligned motion paths for the diagonal case in Figure 17 that minimizes the root mean square position error. The TRMS-PE feature can be calculated from equation 15 of Appendix A.

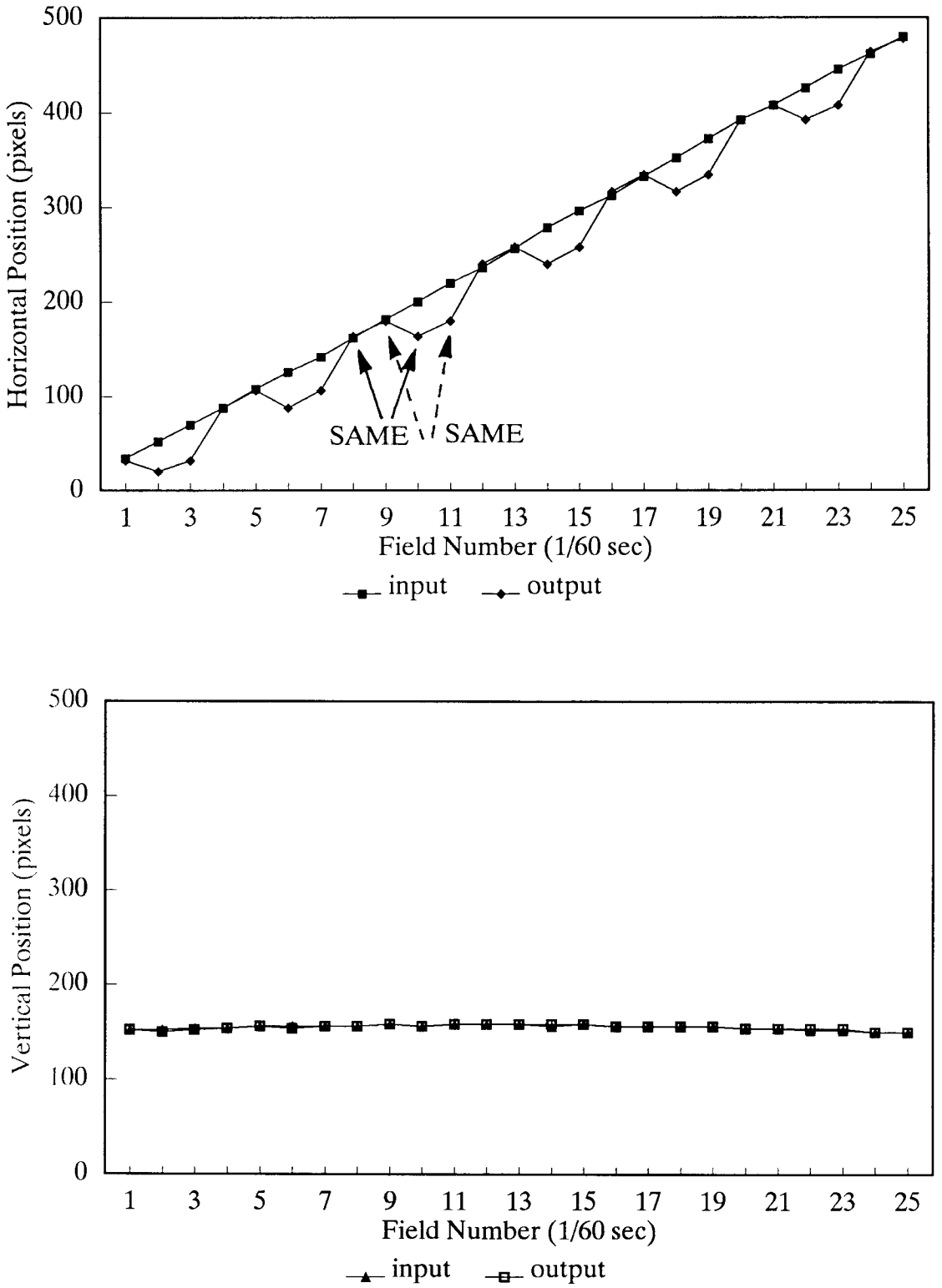

Figure 16. Positions of moving ball as a function of field number for fast motion at the horizontal angle. Ball positions plotted for original NTSC scene and for a code rate of DS1.

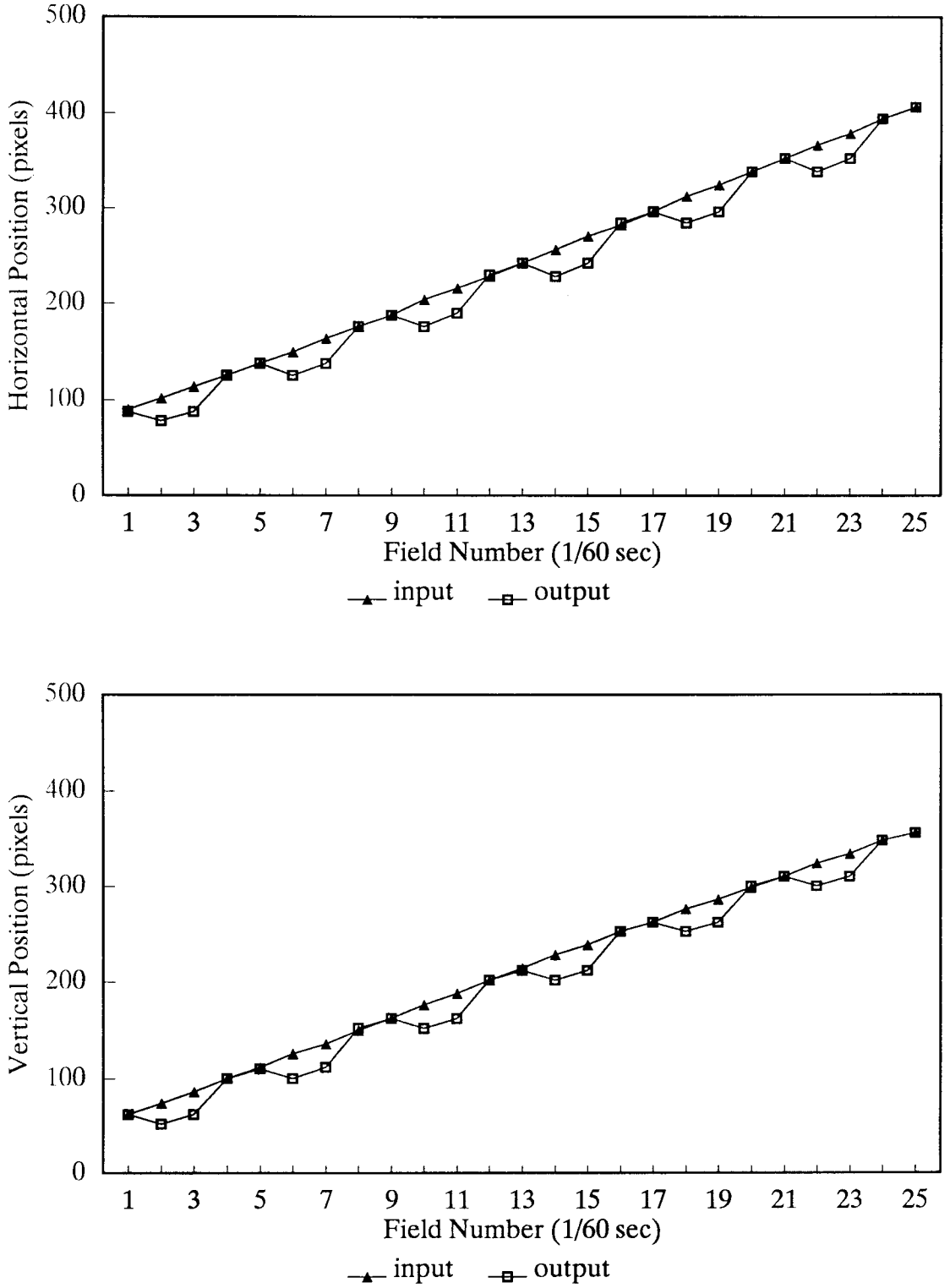

Figure 17. Positions of moving ball as a function of field number for fast motion at the diagonal angle. Ball positions plotted for original NTSC scene and for a code rate of DS1.

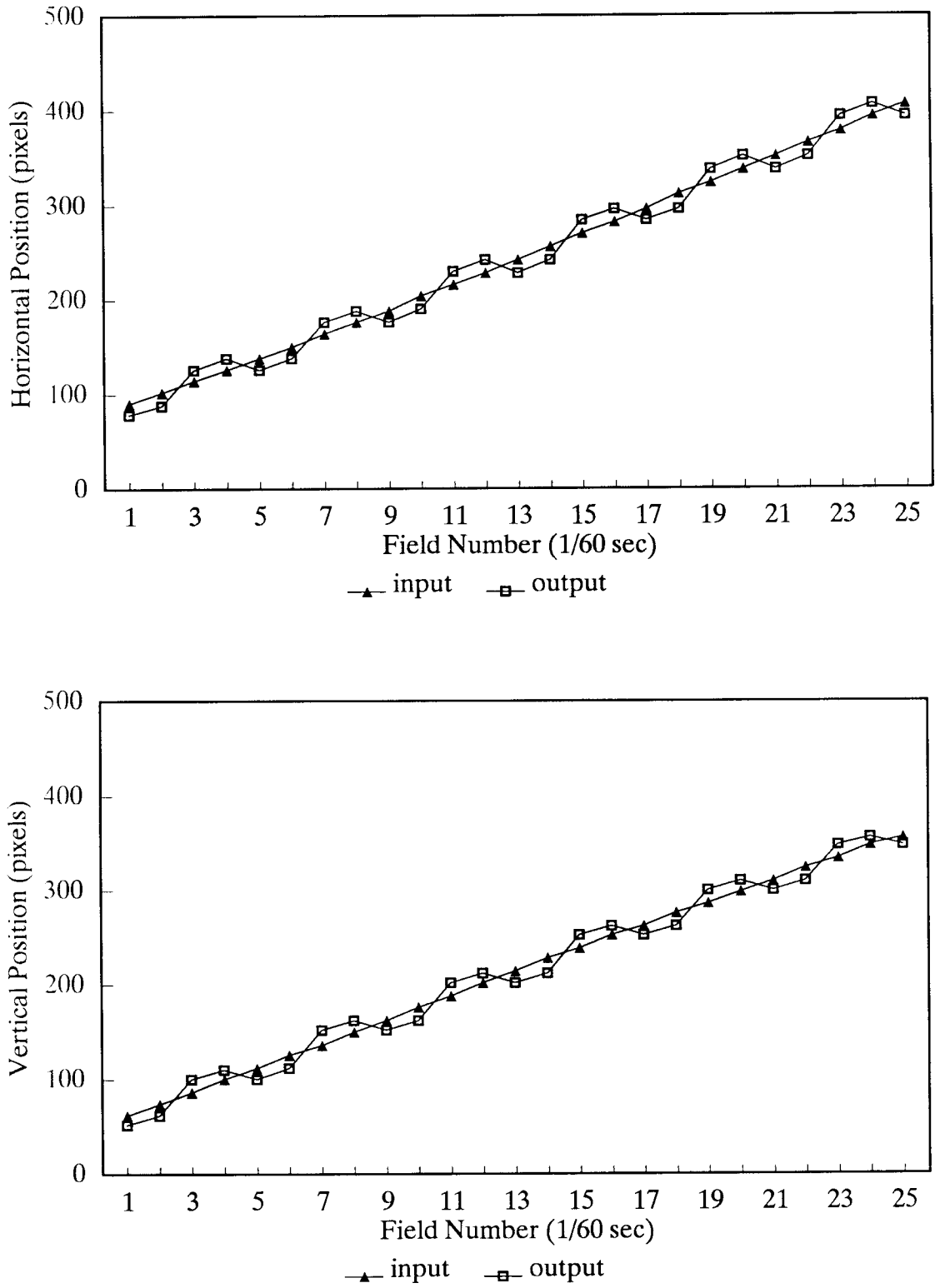

Figure 18. The aligned motion paths for the diagonal case in Figure 17.

The same number of fields in the middle portion of the motion paths was used to calculate the TRMS-PE feature for each trial within each test case. Each input path speed was found by taking the difference in pixels between the endpoints of the middle portion of the motion path divided by the total number of fields (thus, speed is measured in pixels per field). Table 8 summarizes the TRMS-PE results for all the test cases mentioned previously. In Figure 19, the TRMS-PE is plotted verses the speed of the ball for two particular code rates (1/4 DS1, and DS1) and a particular type of motion (horizontal motion). The TRMS-PE feature reflects how far off the output object positions are from the input object positions, on the average. For faster speeds, the output positions are proportionally farther off from the input positions. Thus, the TRMS-PE feature is also proportionally higher. The plot in Figure 19 shows the linear variation of TRMS-PE with speed as described above. In Figure 20, the TRMS-PE is plotted verses the code rate for a particular speed group (fast) and a particular type of motion (diagonal). Here, three trials are shown for each code rate. Since each trial is slightly slower than the previous (three consecutive swings of the ball), there is a slight variation in TRMS-PE between the trials. There is no variation in TRMS-PE with code rate, so the codec is not changing the location to which the output object is placed. The codec is only changing the amount of spatial resolution it allocates to the object (see Figures 14 and 15).

Table 8. Summary Of TRMS-PE Results

Orientation	Code Rate	Speed (Pixels/Field)	TRMS-PE
horizontal	1/4 DS1	17.79	19.13
horizontal	1/4 DS1	17.26	19.01
horizontal	1/4 DS1	16.95	18.13
horizontal	1/4 DS1	9.89	10.84
horizontal	1/4 DS1	9.58	10.86
horizontal	1/4 DS1	9.47	10.51
horizontal	1/4 DS1	6.53	7.91
horizontal	1/4 DS1	6.42	7.51
horizontal	1/4 DS1	6.21	7.08
horizontal	1/2 DS1	17.79	19.09
horizontal	1/2 DS1	17.26	18.88
horizontal	1/2 DS1	16.95	18.13
horizontal	DS1	17.79	19.10
horizontal	DS1	17.26	18.76
horizontal	DS1	16.95	18.01
horizontal	DS1	9.89	10.60
horizontal	DS1	9.58	10.28
horizontal	DS1	9.47	10.15
horizontal	DS1	6.53	7.40
horizontal	DS1	6.42	7.20
horizontal	DS1	6.21	6.70
diagonal	1/4 DS1	17.28	18.49
diagonal	1/4 DS1	16.99	17.93
diagonal	1/4 DS1	16.39	17.71
diagonal	1/2 DS1	17.28	18.54
diagonal	1/2 DS1	16.99	18.19
diagonal	1/2 DS1	16.39	17.52
diagonal	DS1	17.28	18.47
diagonal	DS1	16.99	18.31
diagonal	DS1	16.39	17.88

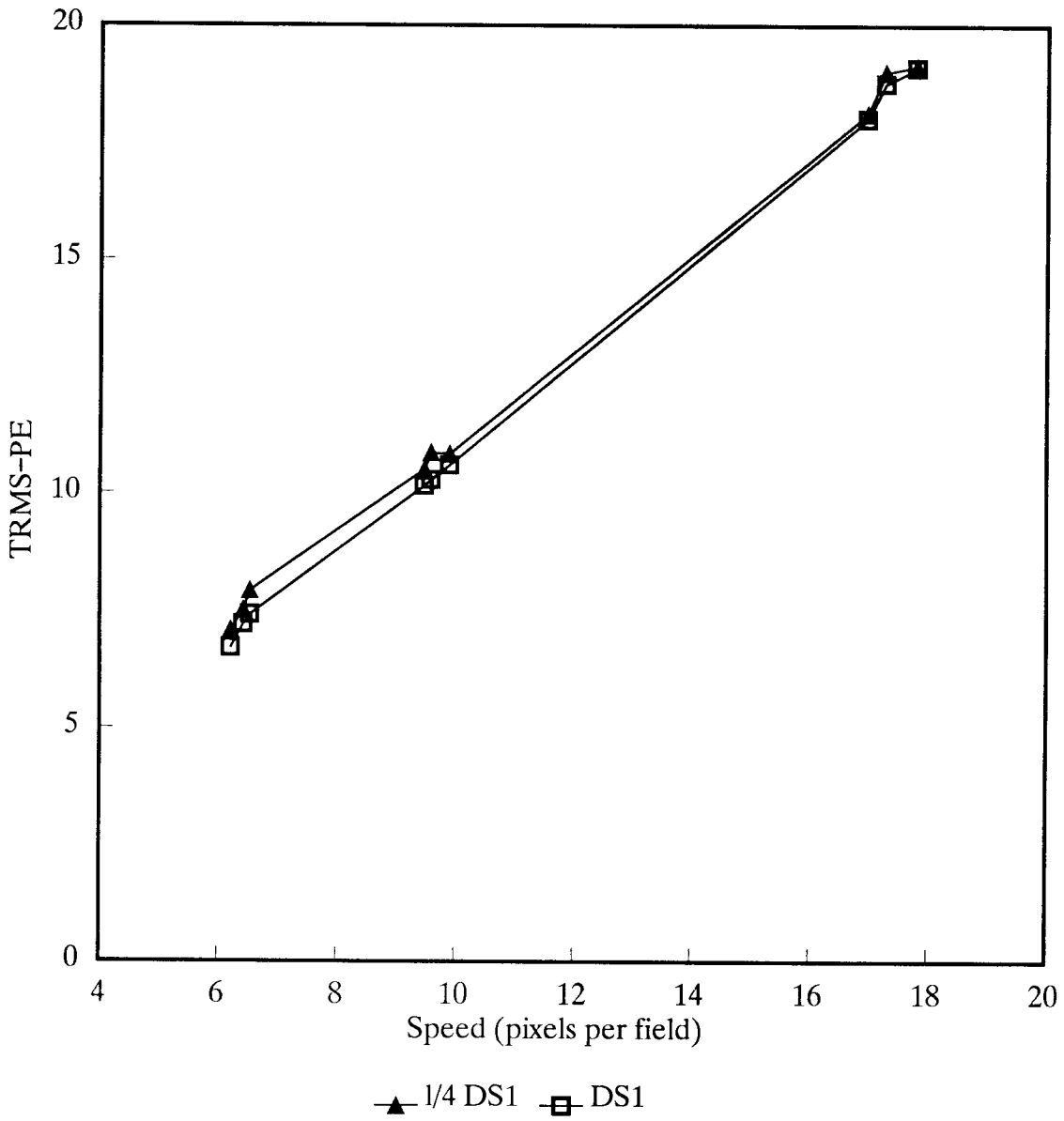

Figure 19. TRMS-PE plotted as a function of horizontal ball speed for code rates of 1/4 DS1 and DS1.

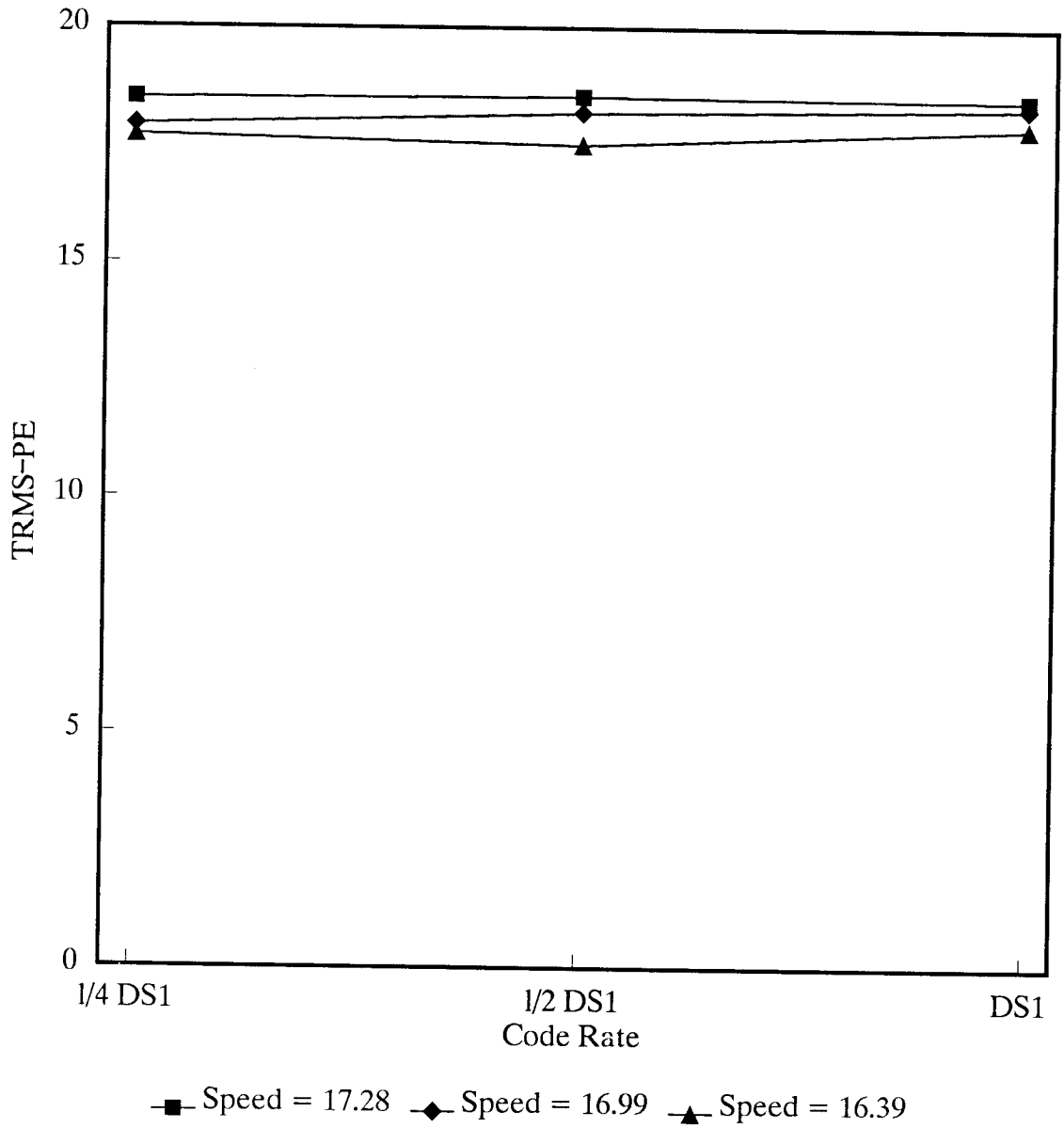

Figure 20. TRMS-PE plotted as a function of code rate for the fast speed group and diagonal motion. Three trials with slightly different speeds are shown.

Applying the TRMS-PE measure of jerkiness to one particular codec has illustrated several important insights into the operation of VTC/VT. First, the TRMS-PE jerkiness measure is very stable; in fact, the small variation in speed between consecutive swings of the ball shows up as a small variation in TRMS-PE, as expected. Second, the jerkiness, or temporal update of the codec, may not vary with code rate (as in Figure 20). The particular codec tested here achieved bit reduction by degrading the spatial resolution of scenes and not the frequency of update. For the particular codec tested, the jerkiness was due to omission of every other frame, regardless of operating bit rate. This simple result would not necessarily have been obtained for codecs that use more sophisticated coding/decoding methods. Other codec algorithms for attaining less jerkiness might trade spatial resolution for more or less temporal positioning accuracy.

2.8 Jerkiness Feature Using Difference Image

The TRMS-PE measure of jerkiness cannot easily be applied to arbitrary video scenes. Section 2.8.1 proposes a measure of jerkiness that can be applied to any video scene. The genesis of this new measure of jerkiness occurred when observing codec input and output video that had been aligned using the single-frame temporal alignment method discussed earlier (as in Figure 14). If one were to compute the difference images of the input and the output video, image pairs that contained no positioning errors (second and third images of each row in Figure 14) would yield smaller difference errors than image pairs that contained positioning errors (first and fourth images of each row in Figure 14). As a function of time, the total composite difference error of a moving object would be composed of two components. One component represents errors due to blurring or distortion of the object. The other component represents errors due to incorrect positioning of the object.

Section 2.8.1 presents a method for extraction of three new features. One of the features will be shown to be intimately related to jerkiness. The other two features represent the average distortion of the output video due to jerkiness and spatial blurring. The exact feature extraction technique and sample VTC/VT results are discussed in detail next.

Applying the TRMS-PE measure of jerkiness to one particular codec has illustrated several important insights into the operation of VTC/VT. First, the TRMS-PE jerkiness measure is very stable; in fact, the small variation in speed between consecutive swings of the ball shows up as a small variation in TRMS-PE, as expected. Second, the jerkiness, or temporal update of the codec, may not vary with code rate (as in Figure 20). The particular codec tested here achieved bit reduction by degrading the spatial resolution of scenes and not the frequency of update. For the particular codec tested, the jerkiness was due to omission of every other frame, regardless of operating bit rate. This simple result would not necessarily have been obtained for codecs that use more sophisticated coding/decoding methods. Other codec algorithms for attaining less jerkiness might trade spatial resolution for more or less temporal positioning accuracy.

2.8 Jerkiness Feature Using Difference Image

The TRMS-PE measure of jerkiness cannot easily be applied to arbitrary video scenes. Section 2.8.1 proposes a measure of jerkiness that can be applied to any video scene. The genesis of this new measure of jerkiness occurred when observing codec input and output video that had been aligned using the single-frame temporal alignment method discussed earlier (as in Figure 14). If one were to compute the difference images of the input and the output video, image pairs that contained no positioning errors (second and third images of each row in Figure 14) would yield smaller difference errors than image pairs that contained positioning errors (first and fourth images of each row in Figure 14). As a function of time, the total composite difference error of a moving object would be composed of two components. One component represents errors due to blurring or distortion of the object. The other component represents errors due to incorrect positioning of the object.

Section 2.8.1 presents a method for extraction of three new features. One of the features will be shown to be intimately related to jerkiness. The other two features represent the average distortion of the output video due to jerkiness and spatial blurring. The exact feature extraction technique and sample VTC/VT results are discussed in detail next.

2.8.1 Feature Extraction Technique

The features are extracted from the undistorted input and distorted output sampled video. The feature extraction technique is computationally efficient and possesses many of the other desirable properties of features that were previously mentioned. The features are extracted from the standard deviation of the difference images (input image minus output image), where the input and output video has been time aligned using the single-frame temporal alignment method. The standard deviation of the difference image is used, instead of the mean or root mean square, because the standard deviation is insensitive to gray level shifts in the sampled video. The exact feature extraction method follows:

1. Video alignment

   Single-frame temporal alignment of the input and output video is performed.

2. Difference image

   For each aligned video image pair, a difference image is formed by subtracting the output image from the input image.

3. Standard deviation of the difference image (SD-DI)

   The standard deviation of each difference image (SD-DI), from step 2 above, is computed as the square root of (the summation of the squares of the image pixel values divided by the total number of pixels, minus the square of the mean of the difference image). Here, the mean of the difference image is computed as the summation of the image pixel values divided by the total number of pixels. See equation 16 in Appendix A for a mathematical definition of SD-DI.

4. Feature computation

   From the time history of SD-DI, from step 3 above, the following three features are computed:

   a. The temporal mean of SD-DI (TM-SD-DI)

      TM-SD-DI is computed as the summation of the SD-DI values divided by the total number of SD-DI values. TM-

SD-DI is primarily related to the average distortion caused by spatial blurring and jerkiness. See equation 17 in Appendix A for a mathematical definition of TM-SD-DI.

b. The temporal standard deviation of SD-DI (TSD-SD-DI)
TSD-SD-DI is computed as the square root of (the summation of the squares of the SD-DI values divided by the total number of SD-DI values, minus the square of TM-SD-DI). This estimate of the standard deviation of the population of SD-DI time samples is asymptotically unbiased for a large number of SD-DI values. An alternate method of computing TSD-SD-DI that is unbiased for a small number of SD-DI values may be used instead (see, for example, Crow et al., 1960). TSD-SD-DI is primarily related to jerkiness. See equation 18 in Appendix A for a mathematical definition of TSD-SD-DI.

c. The temporal root mean square of SD-DI (TRMS-SD-DI)
TRMS-SD-DI is computed as the square root of (the summation of the squares of the SD-DI values divided by the total number of SD-DI values). TRMS-SD-DI is related to the total distortion caused by spatial blurring and jerkiness. See equation 19 in Appendix A for a mathematical definition of TRMS-SD-DI.

To compute the amount of distortion in the difference images with respect to the input images, the SD-DI values could be normalized by the standard deviation of the undistorted input video. Alternatively, normalized features could be obtained by dividing TM-SD-SD, TSD-SD-DI, and TRMS-SD-DI by the temporal mean of the standard deviation of the undistorted input video. Thus, normalized features closer to zero will represent smaller distortions while normalized features closer to one will represent larger distortions.

2.8.2 Sample VTC/VT Results

The VTC/VT imagery of Figure 7 was processed to extract the TM-SD-DI, TSD-SD-DI, and TRMS-SD-DI features as described above. The difference images were obtained by subtracting the output images (rows two, three, and four of Figure 7) from the single-frame temporally aligned input images (row one of Figure 7). Figure 21 shows the resulting difference images, where rows one, two, and three of Figure 7 correspond to codec bit rates of DS1, 1/2 DS1, and 1/4 DS1, respectively. For display purposes only, the difference images of Figure 21 have been scaled such that gray (intensity of 128) represents no error, black (intensities from 0 to 127) represents negative error, and white (intensities from 129 to 255) represents positive error. Note that images 1 and 2 (from left to right) for codec bit rates DS1 and 1/4 DS1 contain small errors, while images 3 and 4 for a codec bit rate of 1/2 DS1 contain small errors. The particular codec under test achieved some of its data compression by discarding every other input frame and repeating every other decoded output frame (one frame represents two images in Figure 7 since the images were grabbed for each field increment of the video recorder, and there are two fields for each NTSC frame). Since the input video was injected asynchronously for each of the three codec bit rates, there was no guarantee that the same frames would be discarded for all bit rates. In Figure 14, the same video frames were discarded for all bit rates (by chance) while in Figure 7 the same frames were discarded for the DS1 and 1/4 DS1 bit rates but different frames were discarded for the 1/2 DS1 bit rate. Thus, care should be taken to process a sufficiently long time sequence of images when extracting the TM-SD-DI, TSD-SD-DI, and TRMS-SD-DI features. Otherwise, inaccurate results may be obtained, particularly for codecs that discard a large number of video frames.

The first two difference images for rate DS1 in Figure 21 contain errors due to blurring. The third and fourth difference images contain errors due to blurring and jerkiness. Examining Figure 7 closely, each image in the NTSC video scene (top row) is unique and shows steady motion of the report crossing the man's face. Meanwhile, the third and fourth codec output images for rate DS1 (second row) are the same as the first and second codec output images, respectively. One can see from Figure 7 that the codec is performing frame repetition. Thus, on the repeated

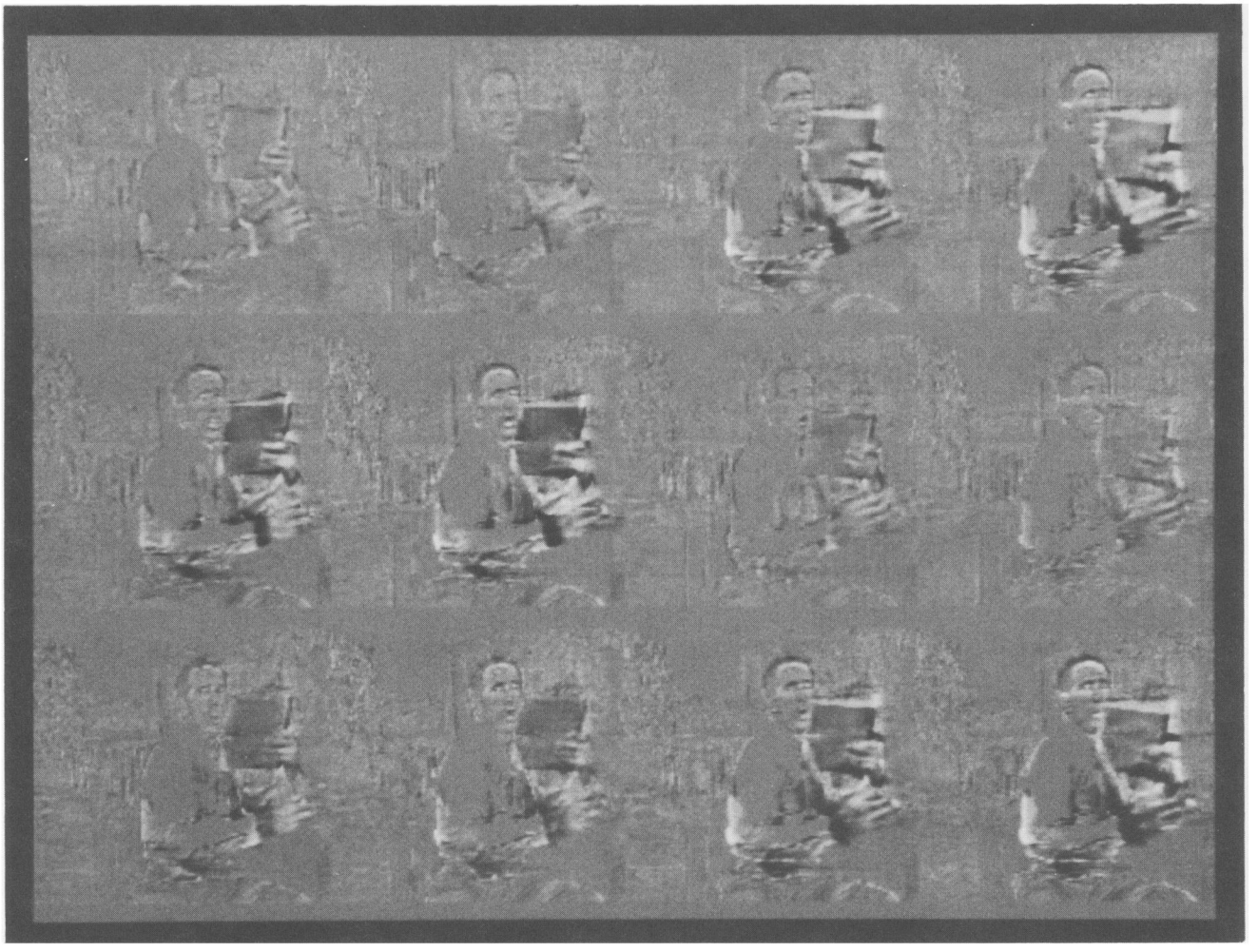

Figure 21. Difference images for VTC/VT imagery of Figure 7. Codec bit rates of DS1 (top row), 1/2 DS1 (second row), and 1/4 DS1 (bottom row).

frame (consisting of images 3 and 4 in Figure 7), large difference errors are obtained and these errors are due to jerkiness in the codec output.

Figure 22 is a graph of the time history of the SD-DI values for the images in Figure 21. The SD-DI values for the first four fields in Figure 22 were calculated from the first four difference images in Figure 21. The ball test scenes (Figures 14 and 15) and the man test scene (Figure 7) were obtained from the same codec. The time history of SD-DI very much resembles the codec output ball position errors (compare with the output ball position error with respect to the true input ball position in Figures 18).

Table 9 presents the computation of the unnormalized TM-SD-DI, TSD-SD-DI, and TRMS-SD-DI features for the eight fields of Figure 22 (for reference, the temporal mean of the standard deviation of the undistorted input video was 77.6). The temporal mean of SD-DI (TM-SD-DI) and the temporal root mean square of SD-DI (TRMS-SD-DI) represents the average and total distortion due to blurring and jerkiness. The temporal standard deviation of SD-DI (TSD-SD-DI) represents the extent of the variation of SD-DI about its mean. More jerky motion will result in larger values of TSD-SD-DI. Curiously, from Table 9, TSD-SD-DI increases slightly with increasing bit rate. This contradicts the earlier TRMS-PE measure of jerkiness which showed that jerkiness was the same for all bit rates (see Figure 20). An explanation for the phenomenon is as follows. The added spatial blurring for low bit rates versus higher bit rates tends to raise the SD-DI curve (increasing TM-SD-DI in Table 9). In the raised SD-DI curve, smaller increases in difference errors due to positioning are obtained, and this results in a flattening of the SD-DI curve (decreasing TSD-SD-DI in Table 9). Subjectively, the TSD-SD-DI measure of jerkiness may be more accurate than the TRMS-PE measure of jerkiness because added spatial blurring tends to reduce the effect of jerkiness. If the object is badly blurred, one cannot tell if the motion is jerky. If the object is focused, one readily notices jerky motion.

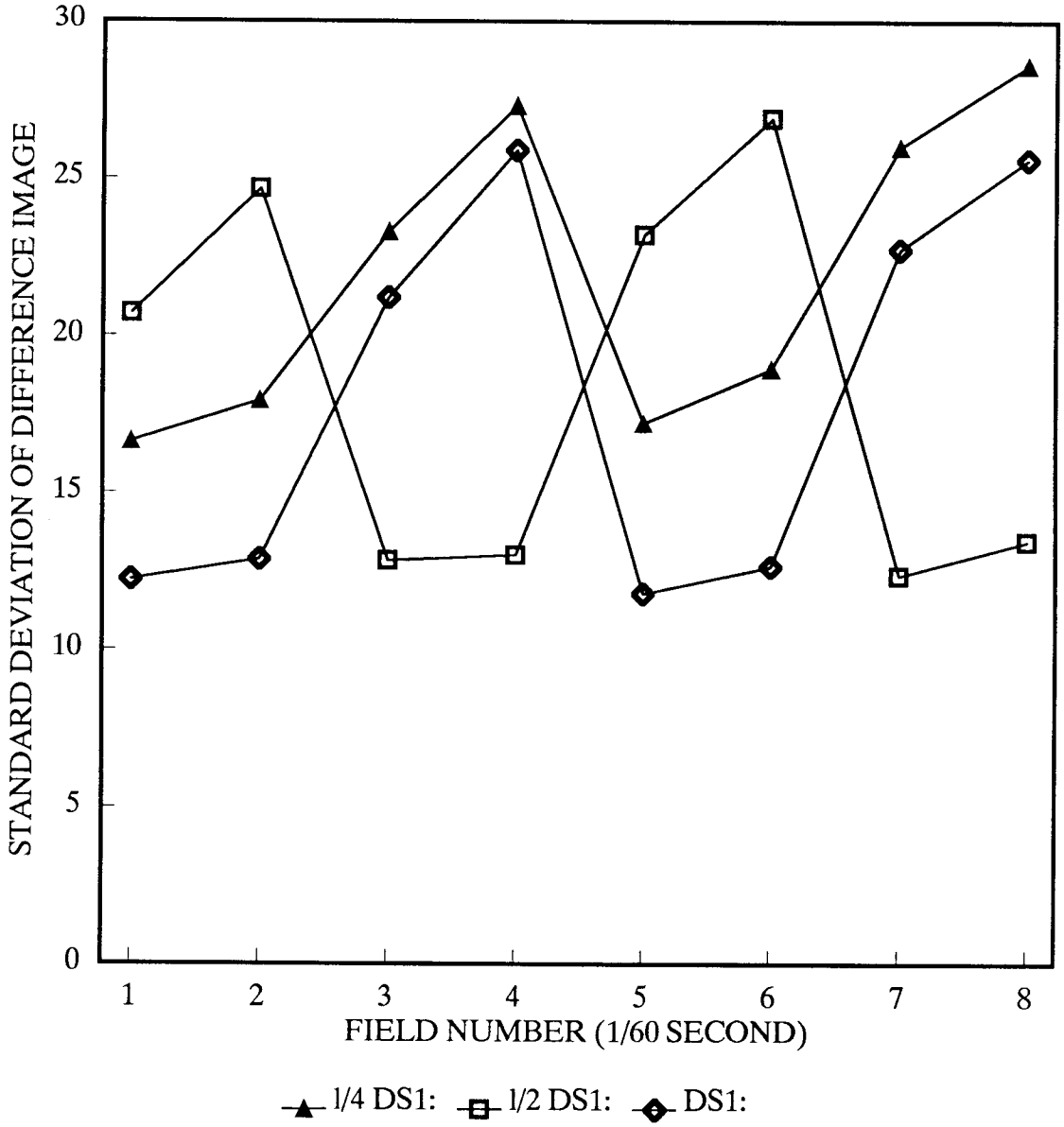

Figure 22. Time history of SD-DI for the difference images of Figure 21. The first four field numbers correspond to the four images in Figure 21.

Table 9. Summary Of SD-DI Features For Figure 22

Scene	TM-SD-DI	TSD-SD-DI	TRMS-SD-DI
DS1	18.1	5.9	19.1
1/2 DS1	18.4	5.7	19.3
1/4 DS1	22.0	4.6	22.5

3. CONCLUSIONS AND RECOMMENDATIONS

Objective feature extraction techniques have been presented that measure the predominant artifacts present in digitally transmitted video systems. Among these artifacts are blurring/smearing, blocking, edge busyness, image persistence, and jerkiness. Features are extracted from the digitized video imagery that reflect degradations perceived by the viewer. The features are sensitive to the type of video being transmitted which is important since the performance of digital codecs depend strongly on the type of video being transmitted. In addition, the features possess many of the desirable properties that humans also possess, including the potential adaptability to focus attention on local disturbances in the video. Thus, the features are expected to correlate strongly with subjective quality ratings.

Depending upon the feature one wishes to extract, the method for temporally aligning the input and distorted output video frames varies. Spatial blurring and jerkiness measures have been presented that do not require the input and output video scenes to be aligned. Other measures, such as edge busyness, blocking, image persistence, and jerkiness for natural motion scenes, require some form of temporal alignment. Two possible methods of temporally aligning the input and output video were presented. The computational requirements of the proposed features varied. However, these computational requirements appear reasonable for modern digital signal processing systems.

Spatial blurring features were presented that relate to the sharpness of the edges in the video imagery. These spatial blurring features appear to be applicable to many types of video imagery, including natural scenes. Blocking, edge busyness, and image persistence were shown to be forms of false edge energy appearing in the output