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

video. Since the importance of edges are well recognized in the areas of human vision and object recognition, it is expected that these spatial blurring, blocking, edge busyness, and image persistence features will correlate especially well with subjective quality ratings.

Two measurement techniques for the jerkiness artifact of digitally transmitted video systems have been proposed. The temporal root mean square position error (TRMS-PE) feature compares the horizontal and vertical positions of a moving object in the output video scene to those in the input video scene. Although a ball was used for the moving object in the presented example, the technique is general enough to substitute any object. The TRMS-PE feature has been shown to be an accurate and repeatable measurement that determines the temporal positioning accuracy of a codec. Additional features were presented that measured the jerkiness of arbitrary video scenes. These included the missing frame ratio (MFR) feature and the TSD-SD-DI feature computed from the standard deviation of the error difference images. Video data compression is often a tradeoff between allocating bits between temporal positioning accuracy and spatial resolution. The ability to measure separately these two attributes raises the possibility of tailoring performance specifications to the application.

Further work needs to be done to determine the optimal method for combining all of the extracted feature values to produce an overall quality rating (the quality classification subsystem shown in Figure 1). Properly combining the many feature measurements into an overall quality assessment rating may require an understanding of the temporal and spatial properties of the eye and brain. To be universally useful, the quality classification subsystem must perform well over a wide range of applications. To obtain this goal, the quality classification system may require user specific application information. Subjective test results on imagery that spans the full range of digitally transmitted video systems should be used to select an optimal set of features, to train the quality classification subsystem, and to evaluate the performance of the completed system.

4. ACKNOWLEDGEMENTS

The author would like to thank Tim W. Butler, Keith E. Junker, Margaret H. Morris, and Dara Parsavand at the Institute For Telecommunication Sciences for writing image processing software that was used to produce images in this report. Further credit and appreciation is extended to Dara Parsavand for his innovative work in developing and testing the temporal root mean square position error (TRMS-PE) feature. Thanks are also extended to Edmund A. Quincy at the Institute For Telecommunication Sciences for the use of video equipment that was required to produce the test data presented in this report.

5. REFERENCES

Barten, P. G. L. (1988), Evaluation of CRT displays with the SQRI method, Society for Information Display International Symposium Digest of Technical Papers, v. XIX, May 24-26, pp. 445-448.

Barten, P. G. L. (1987), The SQRI method: A new method for the evaluation of visible resolution on a display, Proceedings of the Society for Information Display, v. 28/3, pp. 253-262.

Biberman, L. M. (1973), Perception of Displayed Information, Plenum Press, New York-London, pp. 87-119.

Biederman, I. (1985), Human image understanding: recent research and a theory, Computer Vision, Graphics, and Image Processing, v. 32, pp. 29-73.

Carlson, C. R., and R. W. Cohen (1980), A simple psychophysical model for predicting the visibility of displayed information, Proceedings of the Society for Information Display, v. 21/3, pp. 229-246.

CCIR Recommendation 500-3 (1986), Methods for the subjective assessment of the quality of television pictures, CCIR XVth Plenary Assembly, Dubrovnik, v.XI-1, pp. 165-173.

CCIR Recommendation 567-2 (1986), Transmission performance of television circuits designed for use in international connections, International Radio Consultative Committee's Recommendations and Reports of the CCIR, Transmission of Sound Broadcasting and Television Signals Over Long Distances (CMTT), v. XII.

CCIR Recommendation 654 (1986), Subjective quality of television pictures in relation to the main impairments of the analogue composite television signal, International Radio Consultative Committee's Recommendations and Reports of the CCIR, Broadcasting Service (Television), v. XI-1, pp. 223-230.

CCIR Report 313-6 (1986), Assessment of the quality of television pictures, International Radio Consultative Committee's Recommendations and Reports of the CCIR, Broadcasting Service (Television), v. XI-1, pp. 220-223.

CCIR Report 405-5 (1986), Subjective assessment of the quality of television pictures, International Radio Consultative Committee's Recommendations and Reports of the CCIR, Broadcasting Service (Television), v. XI-1, pp. 174-201.

CIE Supplement No. 2 to CIE Publication No. 15 (E-1.3.1) 1971/(TC-1.3.) (1978), Recommendations on uniform color spaces - color difference equations psychometric color terms.

Crow, E.L., Davis, F.A., and M. W. Maxfield (1960), Statistics Manual, Dover Publications Inc., 180 Varick Street, New York, New York.

Fink, D. G. (1975), Electronics Engineers' Handbook, McGraw-Hill Inc., pp. 20-3 to 20-22.

Geuen, W., and H. G. Preuth (1982), New performance criteria of edge detection algorithms, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, v. 3, pp. 1936-1939.

Gonzalez, R. C., and P. Wintz (1987), Digital Image Processing, second edition, Addison-Wesley Publishing Company.

Held, R., Leibowitz, H. W., and H. L. Teuber (1978), Perception, Springer-Verlag Berlin Heidelberg New York, pp. 523-548.

Higgins, G. C. (1977), Image quality criteria, Journal of Applied Photographic Engineering, v. 3, n. 2, pp. 53-60.

Jain, A. K. (1989), Fundamentals of Digital Image Processing, Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Limb, J. O. (1979), Distortion criteria of the human viewer, IEEE Transactions on Systems, Man, and, Cybernetics, v. 9, n. 12, December, pp. 778-793.

Mannos, J. L., and D. J. Sakrison (1974), The effects of a visual fidelity criterion on the encoding of images, IEEE Transactions on Information Theory, v. 20, n. 4, pp. 525-536.

Miyahara, M. (1988), Quality assessments for visual service, IEEE Communications Magazine, October, pp. 51-60.

Meiseles, H. (1988), Objective measurement methods of motion artifacts for 45 Mbit, NTSC, DPCM, bit-reduction video codecs, 130th SMPTE Technical Conference, October 15-19.

Miyahara, M., and Y. Yoshida (1988), Mathematical transform of (R, G, B) color data to Munsell (H, V, C) color data, SPIE Visual Communications and Image Processing, v. 1001, pp. 650-657.

Murakami, H., Hashimoto, H., and Y. Hatori (1988), Quality of band-compressed TV services, IEEE Communications Magazine, October, pp. 61-69.

Nesenbergs, M. (1989), Image data compression overview: issues and partial solutions, U.S. Department of Commerce, National Telecommunications and Information Administration, NTIA Report 89-252, October.

Newhall, S. M., Nickerson, D., and D. B. Judd (1943), Final report of the O.S.A. Subcommittee on the spacing of the Munsell colors, Journal of the Optical Society of America, v. 33, n. 7, pp. 385-418.

Ohtsuka, S., Inoue, M., and K. Watanabe (1988), Quality evaluation of pictures with multiple impairments based on visually weighted error, Society for Information Display International Symposium Digest of Technical Papers, v. XIX, May 24-26, pp. 428-431.

Owens, R., Venkatesh, S., and J. Ross (1989), Edge detection is a projection, Pattern Recognition Letters, v. 9, pp. 233-244.

Oppenheim, A. V., and R. W. Schafer (1975), Digital Signal Processing, Prentice-Hall, Englewood Cliffs, New Jersey.

Pearson, D. E. (1980), A three-stage process for the evaluation of image quality, Proceedings of the Society for Information Display, v. 21/3, pp. 271-278.

Sakrison, D. J. (1977), On the role of the observer and a distortion measure in image transmission, IEEE Transactions on Communications, v. 25, n. 11, November, pp. 1251-1267.

Shapley, R. M., and D. J. Tolhurst (1973), Edge detectors in human vision, J. Physiol., v. 229, pp. 165-183.

Taylor, J. M., Murch, G. M., and P. A. McManus (1989), TekHVC™: A uniform perceptual color system for display users, Proceedings of the Society for Information Display, v. 30/1, pp. 15-21.

Task, H. L. (1979), An evaluation and comparison of several measures of image quality, Aerospace Medical Research Lab Wright-Patterson AFB OH, AMRL-TR-79-7, January (available on Micro-fische, AD-A069 690).

Task, H. L., Pinkus, A. R., and J. P. Hornseth (1978), A comparison of several television display image quality measures, Proceeding of the Society for Information Display, v. 19/3, pp. 113-119.

Toit, T. C. du, and J. G. Lourens (1988), Automatic detection of image impairments, IEEE International Conference on Acoustics, Speech, and Signal Processing, Multidimensional Signal Processing, v. 2, April 11-14, pp. 1080-1083.

Tzafestas, S. G. (1986), Multidimensional Systems: Techniques and Applications, Marcel Dekker Inc, New York and Basel.

Tomich, D. J., Quincy, E. A., and D. Parsavand, (1989), Expert pattern recognition assessment of image quality, IEEE International Conference on Acoustics, Speech, and Signal Processing, Multidimensional Signal Processing, Audio & Electroacoustics, v. 3, May 23-26, pp. 1799-1802.

Westernik, J. H. D. M., and J. A. J. Roufs (1988), A local basis for perceptually relevant resolution measures, Society for Information Display International Symposium Digest of Technical Papers, v. XIX, May 24-26, pp. 360-363.

6. BIBLIOGRAPHY

Allnatt, J. W. (1983), Transmitted-Picture Assessment, John Wiley & Sons, New York.

Beaton, R. J. (1988), Linear systems metrics of image quality for flat-panel displays, SPIE Image Processing, Analysis, Measurement, and Quality, v. 901, pp. 144-151.

Goodman, J. S., and D. E. Pearson (1979), Multidimensional scaling of multiply-impaired television pictures, IEEE Transactions on Systems, Man, and Cybernetics, v. 9, n. 6, June, pp. 353-356.

Lewis, N. W., and J. W. Allnatt (1965), Subjective quality of television pictures with multiple impairments, Electronic Letters, v. 1, n. 7, September, pp. 187-188.

Lourens, J. G., and M. W. Coetzer (1987), Image impairment detection using digital signal processing, International Conference of Television Measurements, 3rd, pp. 33-38.

Moon, D. L. (1988), Image quality for discrete-element displays: variables, metrics, and measurements, SPIE Image Processing, Analysis, Measurement, and Quality, v. 901, pp. 161-170.

Netravali, A. N., and B. G. Haskell (1988), Digital Pictures Representation and Compression, Plenum Publishing Corporation, New York, New York, pp. 245-297.

Pratt, W. K. (1978), Digital Image Processing, John Wiley & Sons, New York, pp. 162-200.

Smith-Kerker, P. L., and R. G. Bias (1988), The effect of color encoding on the subjective quality of color images, Society for Information Display International Symposium Digest of Technical Papers, v. XIX, May 24-26, pp. 85-88.

7. APPENDIX A: EQUATIONS

This appendix describes all of the equations that were used to compute the feature values in this report.

Equation (1): Single-frame Temporal Alignment

Let $i(v,h,t_i)$ be the digitized input video sequence where v is the vertical sampling index, h is the horizontal sampling index, and t_i is the input frame sampling index. Here $v = \{1, 2, ..., N_v\}$, and $h = \{1, 2, ..., N_h\}$, where N_v is the total number of vertical pixels, and N_h is the total number of horizontal pixels. Similarly, let the digitized output video sequence be represented by $o(v,h,t_o)$, where t_o is the output frame sampling index. Assume that the input video sequence and the output video sequence are sampled at the same frame rate. Then, given an output reference frame $t_o = r$, the single-frame temporal alignment problem is to find the closest corresponding input frame $t_i = m$. The single-frame temporal alignment proposed here assumes that a priori knowledge is available which gives the range of the closest corresponding input frame index (say, from $t_i = t_l$ to t_u, where t_l and t_u are the lower and upper limits, respectively) and that the input video sequence contained moving and/or changing scenes. Then the closest matching input frame $t_i = m$ can be found as the t_i that minimizes the standard deviation of the error (accumulated over all pixels in the error image) or

$$\sqrt{\left\{\frac{1}{N_v N_h}\sum_{v=1}^{N_v}\sum_{h=1}^{N_h}\left[i(v,h,t_i)-o(v,h,r)\right]^2\right\} - \left\{\frac{1}{N_v N_h}\sum_{v=1}^{N_v}\sum_{h=1}^{N_h}\left[i(v,h,t_i)-o(v,h,r)\right]\right\}^2}$$

where t_i falls within the range from t_l to t_u, inclusive. Pairing of the input and output video frame indices is performed as ..., ($t_i = m-1$, $t_o = r-1$), ($t_i = m$, $t_o = r$), ($t_i = m+1$, $t_o = r+1$), ...

Equation (2): Missing Frame Ratio (MFR)

To apply multi-frame temporal alignment to the output video frame sequence $o(v,h,t_o)$, where $t_o = \{1, 2, ..., N_o\}$, and N_o is the total number

of output video frames, the method of finding the closest matching input video frame (equation 1, Appendix A) is applied to every output video frame $t_o = \{1, 2, ..., N_o\}$. The computation of the closest input video frame to output video frame $t_o = 1$, may be used to refine the estimates of the lower (t_l) and upper (t_u) input frame limits for output video frame $t_o = 2$. While performing multi-frame alignment, the closest input video frame index to each one of the output video frames is stored. Let the number of <u>unique</u> input frame indices within the set of stored indices be N_u. N_u will be less than N_o if input video frames have been omitted in the output. Then, the missing frame ratio (MFR) is calculated as

$$MFR = \frac{N_o - N_u}{N_o}$$

<u>Equation (3): Mean of the Sobel Image (M-SI)</u>

If the Sobel edge extracted image is given by s(v,h), where v and h represent the vertical and horizontal sampling indices of the video image, then the mean of s(v,h) over a given sub-regional area is given by

$$M\text{-}SI = \frac{1}{N_A}\sum_{v}\sum_{h} s(v,h)$$

where the summation is performed over the sub-regional area and N_A is the total number of pixels within the sub-regional area.

<u>Equation (4): Standard Deviation of the SI (SD-SI)</u>

Following the notation established for equation (3) in Appendix A, the standard deviation of the sub-regional Sobel image is

$$SD\text{-}SI = \sqrt{\{\frac{1}{N_A}\sum_v \sum_h s(v,h)^2\} - (M\text{-}SI)^2}$$

Equation (5): Root Mean Square of the SI (RMS-SI)

Following the notation established for equation (3) in Appendix A, the root mean square of the sub-regional Sobel image is

$$RMS\text{-}SI = \sqrt{\frac{1}{N_A}\sum_v \sum_h s(v,h)^2}$$

Equation (6): Number of Pixels Greater than Threshold of SI (NPGT-SI)

Following the notation established in equation (3) in Appendix A and letting T be the chosen threshold, the number of pixels greater than T within the sub-regional Sobel image is

$$NPGT\text{-}SI = \sum_v \sum_h u(v,h)$$

$$\textit{where}$$

$$u(v,h) = 1 \ \textit{if} \ s(v,h) > T$$
$$= 0 \ \textit{otherwise}$$

Equation (7): Mean of the Positive Sobel Difference Image (M-PSDI)

If the Sobel difference image (Sobel filtered input image minus the Sobel filtered output image) is given by $s_d(v,h)$, where v and h represent the vertical and horizontal sampling indices, then the mean of the sub-regional, positive part of the sobel difference image is

$$M\text{-}PSDI = \frac{1}{N_A}\sum_v \sum_h s_d(v,h) \quad , \; s_d(v,h)>0$$

where N_A is the total number of points within the sub-regional area.

Equation (8): Standard Deviation of the PSDI (SD-PSDI)

Following the notation of equation (7) in Appendix A, the standard deviation of the sub-regional, positive part of the Sobel difference image is

$$SD\text{-}PSDI = \sqrt{\frac{1}{N_A}\sum_v \sum_h s_d^2(v,h) - (M\text{-}PSDI)^2} \quad , \; s_d(v,h)>0$$

Equation (9): Root Mean Square of the PSDI (RMS-PSDI)

Following the notation of equation (7) in Appendix A, the root mean square of the sub-regional, positive part of the Sobel difference image is

$$RMS\text{-}PSDI = \sqrt{\frac{1}{N_A}\sum_v \sum_h s_d^2(v,h)} \quad , \; s_d(v,h)>0$$

Equation (10): Number of Pixels Greater than Threshold of PSDI (NPGT-PSDI)

Following the notation of equation (7) in Appendix A, and letting T_p be the chosen threshold, the number of pixels greater than T_p within the sub-regional, positive part of the Sobel difference image is

$$NPGT\text{-}PSDI = \sum_v \sum_h u_p(v,h)$$

where

$$u_p(v,h) = 1 \ \ if \ s_d(v,h) > T_p$$
$$= 0 \ \ otherwise$$

Equation (11): Mean of the Negative Sobel Difference Image (M-NSDI)

If the Sobel difference image (Sobel filtered input image minus the Sobel filtered output image) is given by $s_d(v,h)$, where v and h represent the vertical and horizontal sampling indices, then the mean of the sub-regional, negative part of the sobel difference image is

$$M\text{-}NSDI = \frac{1}{N_A} \sum_v \sum_h s_d(v,h) \quad , \ s_d(v,h) < 0$$

where N_A is the total number of points within the sub-regional area.

Equation (12): Standard Deviation of the NSDI (SD-NSDI)

Following the notation of equation (11) in Appendix A, the standard deviation of the sub-regional, negative part of the Sobel difference image is

$$SD\text{-}NSDI = \sqrt{\frac{1}{N_A} \sum_v \sum_h s_d^2(v,h) - (M\text{-}NSDI)^2} \quad , \ s_d(v,h) < 0$$

Equation (13): Root Mean Square of the NSDI (RMS-NSDI)

Following the notation of equation (11) in Appendix A, the root mean square of the sub-regional, negative part of the Sobel difference image is

$$RMS\text{-}NSDI = \sqrt{\frac{1}{N_A}\sum_v \sum_h s_d^2(v,h)} \quad , \; s_d(v,h)<0$$

Equation (14): Number of Pixels Less than Threshold of NSDI (NPLT-NSDI)

Following the notation of equation (11) in Appendix A, and letting T_n be the chosen threshold, the number of pixels less than T_n within the sub-regional, negative part of the Sobel difference image is

$$NPLT\text{-}NSDI = \sum_v \sum_h u_n(v,h)$$

$$where$$

$$\begin{aligned} u_n(v,h) &= 1 \;\; if \;\; s_d(v,h)<T_n \\ &= 0 \;\; otherwise \end{aligned}$$

Equation (15): Temporal Root Mean Square Position Error (TRMS-PE)

Let the input, or reference, vertical and horizontal positions of the moving object be represented by $v_i(t_i)$ and $h_i(t_i)$, where t_i represents the frame sampling index such that $t_i = \{1, 2, 3, ..., N_i\}$, and N_i is the total number of time samples for the input object path. Similarly, let the output vertical and horizontal positions of the moving object be represented by $v_o(t_o)$ and $h_o(t_o)$, where $t_o = \{1, 2, 3, ..., N_o\}$, and N_o is the total number of time samples for the output object path. In order to measure TRMS-PE, the input and output motion paths have to be aligned to compensate for the absolute video delay of the device under test. The alignment procedure described here corresponds to what a viewer would observe if that viewer was insensitive to the absolute video delay. Assume that the output motion path corresponds to some portion of the input path, and is thus contained completely within the input motion path (i.e., $N_o < N_i$). Then, the TRMS-PE feature is computed as the minimum root mean square position error of the output motion path with respect to the input motion path, where the minimization is performed over all possible time shifts $s = \{0, 1, 2, ..., N_i\text{-}N_o\}$ of the two motion paths. In equation form, the computation may be written as

$$TRMS\text{-}PE = \underset{s}{MIN} \sqrt{\frac{1}{N_o} \sum_{t_o=1}^{t_o=N_o} [v_i(t_o+s) - v_o(t_o)]^2 + [h_i(t_o+s) - h_o(t_o)]^2}$$

As in equation (1) in Appendix A, a priori knowledge of the absolute video delay may be used to narrow the range of time shifts.

Equation (16): Standard Deviation of Difference Image (SD-DI)

Assume that the input and output video sequences have been aligned using single frame temporal alignment given in equation (1) in Appendix A, and thus each output video frame has been paired with some input video frame. Let each pair of video frames be represented by the index p = {1, 2, 3, ..., N_p), where N_p is the total number of input/output pairs. Let the difference image (input image minus output image) of each input/output pair be represented by $d_p(v,h)$, where v and h are the vertical and horizontal sampling indices. Then, the standard deviation of the difference image over a sub-regional area is given by

$$(SD\text{-}DI)_p = \sqrt{\{\frac{1}{N_A}\sum_v \sum_h d_p^2(v,h)\} - \{\frac{1}{N_A}\sum_v \sum_h d_p(v,h)\}^2}$$

where N_A is the total number of points within the sub-regional area.

Equation (17): Temporal Mean of SD-DI (TM-SD-DI)

Following the notation of equation (16) in Appendix A, the temporal mean of the time history of SD-DI is given as

$$TM\text{-}SD\text{-}DI = \frac{1}{N_p} \sum_{p=1}^{p=N_p} (SD\text{-}DI)_p$$

Equation (18): Temporal Standard Deviation of SD-DI (TSD-SD-DI)

Following the notation of equation (16) and (17) in Appendix A, the temporal standard deviation of the time history of SD-DI is given as

$$TSD\text{-}SD\text{-}DI = \sqrt{\left\{ \frac{1}{N_p} \sum_{p=1}^{p=N_p} (SD\text{-}DI)_p^{\,2} \right\} - (TM\text{-}SD\text{-}DI)^2}$$

Equation (19): Temporal Root Mean Square of SD-DI (TRMS-SD-DI)

Following the notation of equation (16) in Appendix A, the temporal root mean square of the time history of SD-DI is given as

$$TRMS\text{-}SD\text{-}DI = \sqrt{\frac{1}{N_p} \sum_{p=1}^{p=N_p} (SD\text{-}DI)_p^{\,2}}$$

8. APPENDIX B: FILTERS

This appendix describes the median filter that was used to precondition the digitized video images and the Sobel filter that was used to extract edges from the preconditioned images.

Median Filter

A median filter will remove noise spikes from the image without significantly blurring the edges. Very fine detail, such as sharp corners, will be removed by the median filter. For this report, a 3 x 3 median filter was used. Figure B-1 shows 9 image pixel values (X_1, X_2, ..., X_9) within the 3 x 3 filter window. In Figure B-1, the 3 x 3 filter window is centered on the image pixel value X_5.

X_1	X_2	X_3
X_4	X_5	X_6
X_7	X_8	X_9

Figure B-1. Filter window centered on image pixel value X_5.

The median filter outputs the image pixel value that is the median of the 9 image pixel values (X_1, X_2, ..., X_9). That is, the 9 pixel values are first sorted from low to high, and then the middle value is selected as the median. The median filtered image is obtained by sliding the 3 x 3 window over the entire input image. At each pixel for which the mask is centered in the input image, the median value is placed in the output image. Note that as an edge is crossed, one side or the other dominates the window and the output switches sharply.

Sobel Filter

The Sobel filter is an edge extraction filter that is implemented using two filters. One filter is designed to extract horizontal edges from the image and the other filter is designed to extract vertical edges. The outputs from the two filtering operations are then combined to give a composite edge extracted image. Figure B-2 gives the filter

mask that extracts the horizontal edges. Figure B-3 gives the filter mask that extracts the vertical edges.

-1	-2	-1
0	0	0
1	2	1

Figure B-2. Horizontal edge extraction filter mask.

-1	0	1
-2	0	2
-1	0	1

Figure B-3. Vertical edge extraction filter mask.

If both of the masks shown in Figures B-2 and B-3 are centered on pixel value X_5, as in Figure B-1, then the output response at pixel location X_5 from the horizontal edge extraction filter is

$$G_h = -1*X_1 - 2*X_2 - 1*X_3 + 1*X_7 + 2*X_8 + 1*X_9$$

and the output response at pixel location X_5 from the vertical edge extraction filter is

$$G_v = -1*X_1 + 1*X_3 - 2*X_4 + 2*X_6 - 1*X_7 + 1*X_9$$

Note that horizontal edges result in an output response G_h and vertical edges result in an output response G_v. Diagonal edges result in output responses G_h and G_v. The composite output response at pixel location X_5 from both filters is computed as

$$G = [G_h^2 + G_v^2]^{1/2}$$

The output image pixel values are obtained by computing the filter response G at each corresponding pixel in the input image, where both filter masks (Figure B-2 and B-3) are centered on the input image pixel.