Algorithm for automated eye strain reduction in real stereoscopic images and sequences

J. S. McVeigh[1], M. W. Siegel[2] and A. G. Jordan[1]

[1]Department of Electrical and Computer Engineering
[2]Robotics Institute, School of Computer Science
Carnegie Mellon University, Pittsburgh, PA 15213

ABSTRACT

Eye strain is often experienced when viewing a stereoscopic image pair on a flat display device (e.g., a computer monitor). Violations of two relationships that contribute to this eye strain are: 1) the accommodation/convergence breakdown and 2) the conflict between interposition and disparity depth cues. We describe a simple algorithm that reduces eye strain through horizontal image translation and corresponding image cropping, based on a statistical description of the estimated disparity within a stereoscopic image pair. The desired amount of translation is based on the given stereoscopic image pair, and, therefore, requires no user intervention.

In this paper, we first develop a statistical model of the estimated disparity that incorporates the possibility of erroneous estimates. An estimate of the actual disparity range is obtained by thresholding the disparity histogram to avoid the contribution of false disparity values. Based on the estimated disparity range, the image pair is translated to force all points to lie on, or behind, the screen surface. This algorithm has been applied to diverse real stereoscopic images and sequences. Stereoscopic image pairs, which were often characterized as producing eye strain and confusion, produced comfortable stereoscopy after the automated translation.

Keywords: stereoscopic image processing, vision-based algorithms for image enhancement, depth cues.

1. INTRODUCTION

A stereoscopic image pair is defined as the combination of left- and right-eye images obtained by two horizontally displaced cameras. By presenting the appropriate view to each eye, the viewer experiences the sensation of stereopsis, which provides added realism through improved depth perception. However, when stereoscopic imagery is displayed on a two-dimensional device, the viewer often experiences eye-strain and confusion. This eye-strain is most likely due to the violation of relationships that exist in the real-world.

When a viewer observes an object in the real-world, the eyes converge to a specific point, reducing retinal disparity, and focus (accommodate) to the same point. Conversely, when viewing a stereo-pair on a flat screen, the eyes converge as if portions of the image are located at different distances, but they remain focused on the plane of the screen. This breakdown between the normal accommodation/convergence relationship has been shown to be a major cause of eye-strain.[6,8,10,15]

Another conflict between depth cues arises when an object in the stereo-pair appears to lie in front of the screen, based on the object's disparity, but a portion of the object is clipped by the screen (or image window) surround. The interposition depth cue indicates that the image surround is in front of the object, which is in direct opposition to the disparity depth cue. This conflict causes depth ambiguity and confusion.

We would like a method to avoid, or at least minimize, the breakdown of the accommodation/convergence relationship and the conflict between interposition and disparity depth cues to reduce eye-strain and confusion. Miyashita and Uchida attempt to eliminate the mismatch between accommodation and convergence through the design of a special optical system, which uses an aperture stop to increase the focal depth.[8] This solution requires additional optical hardware and possibly another optical system to compensate for discomfort introduced by the aperture stop. Pastoor suggests reducing the disparity range through an adjustment of the stereo convergence to minimize the mismatch.[10] However, this solution effectively limits the depth resolution of the stereoscopic imagery.

We can achieve our goal of reducing eye-strain without additional optical hardware or a reduction in the disparity range by forcing all image points to lie on, or behind, the screen surface (i.e., have non-negative disparity). This is accomplished through a simple horizontal translation (and appropriate cropping) of one image with respect to the other by the minimum disparity value of the given stereo-pair. This technique preserves the disparity range and depth ordering, reduces the imbalance between accommodation and convergence, and completely eliminates the conflict between the interposition and disparity depth cues. While, after translation, the imagery may not be the geometrically correct perspective view (as described by Grinberg, *et. al.*[5]), we suggest that the geometrically correct view is not necessarily the best view if it introduces eye-strain or confusion.

Lipton describes such a methodology for the generation of stereoscopic computer graphics, where the synthetic imagery is positioned to lie behind the screen surface.[6] A similar approach for computer graphics is taken by Ware, *et. al.*, where the z-buffer of the graphics system displaying the stereoscopic imagery is sampled to find the disparity range.[14] The task of calculating the disparity range is considerably more difficult for real stereoscopic image pairs; the possibility of erroneous estimates must be handled appropriately to avoid an inaccurate range calculation. Although we could easily specify an over-cautious disparity range to ensure that all points lie behind the screen surface, in doing so we might increase the mismatch between the accommodation/convergence relationship. Therefore, in this paper, we attempt to accurately calculate the true disparity range by taking into consideration the cause and effect of erroneous disparity estimates.

The contributions of this work are: 1) the development of a statistical model for false disparity estimates, 2) the incorporation of this model for the accurate estimation of the actual disparity range from the disparity histogram, and 3) the application of this range estimate for eye-strain reduction through horizontal image translation. In the next section, we detail the process of disparity estimation and the causes of estimation errors. We follow this with the development of a statistical model of the estimated disparity values and provide our method for selecting the actual range. We then illustrate the usefulness of this algorithm on two real stereoscopic sequences. Finally, we conclude with areas for future work, including controlled human factors experiments to substantiate our informal results.

2. DISPARITY ESTIMATION

Throughout this paper, we will assume that the stereoscopic imagery is obtained from a pair of cameras with parallel, coplanar, and horizontally displaced axes. A point in the scene, P, visible to both cameras (i.e., unoccluded) will map to the corresponding points $\boldsymbol{p}_l$ and $\boldsymbol{p}_r$ in the left- and right-eye images, respectively. Disparity is the vector distance, in pixels, between corresponding points within appropriately frames images, and is related to the distance of the scene point from the stereo camera. Due to the parallel camera configuration, the disparity will be only in the horizontal direction. The relationship between the corresponding points is given by,

$$\boldsymbol{p}_l = \begin{bmatrix} x_l \\ y_l \end{bmatrix} = \begin{bmatrix} x_r \\ y_r \end{bmatrix} - \begin{bmatrix} d(x_r, y_r) \\ 0 \end{bmatrix} \tag{1}$$

where $d(x_r, y_r)$ is the disparity from the right image pixel to the corresponding left image pixel. A disparity value for a point visible in only one of the views (i.e., occluded) is, obviously, undefined.

Figure 1a depicts the relationship between disparity and the perceived range of objects within a stereo-pair when viewed on a display device. The viewer's eyes are separated by the interocular separation (e), and are located in front of the screen surface (plane of projection) by a distance s. The angles subtended by the two eyes represent the fields of view of each eye through the display's stereo-window. Three points are included in the figure at the range they are perceived to lie based on their disparities, and projective rays passing through each eye and point are drawn. The intersection of these rays with the plane of projection are the pixel locations of the corresponding points in the left- and right-eye images (l_i and r_i, where $i = 1, 2, 3$).

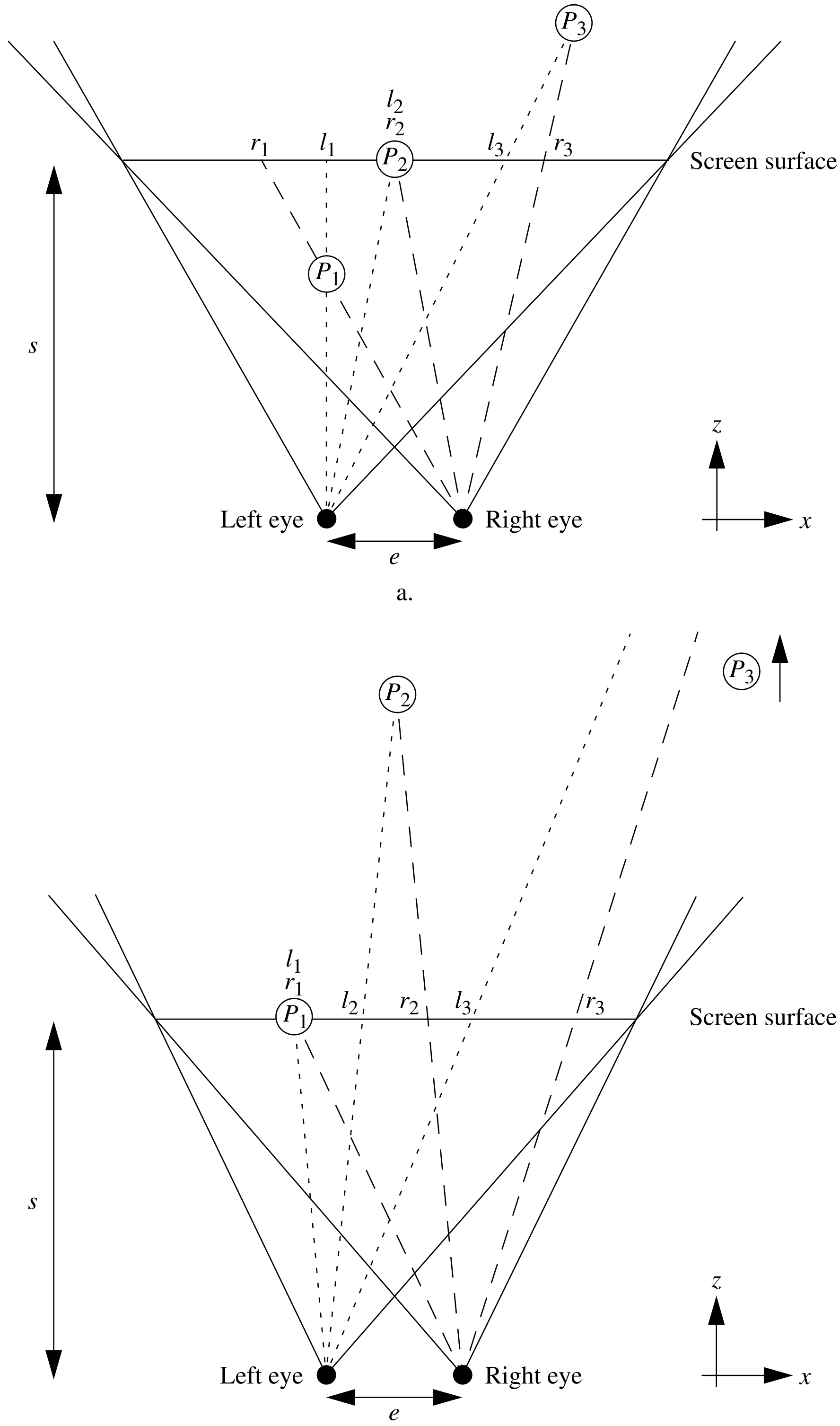


Figure 1: Relationship between disparity and perceived depth: a) original scene, b) scene after horizontal translation and cropping.

The three points depicted in the diagram represent the three possible situations for disparity in a stereoscopic image. The first point (P_1) effectively lies between the viewer and the screen surface (negative disparity); the second point (P_2) lies on the screen surface and has no, or zero, disparity; and the third point (P_3) lies behind the screen surface (positive disparity). The relationship between the point's on screen disparity and its perceived range (z) is given by,

$$z_i = s\left(\frac{e}{e - d_i}\right) \tag{2}$$

where $d_i = r_i - l_i$ is measured in a physical distance unit.

If the disparity values of these points are known, the right image can be translated to the right by the absolute value of d_1 to force all three points to lie on, or behind, the screen surface. The left-border of the left image and the right-border of the right image also would need to be cropped by the translation distance to ensure that the resulting images cover the same field of view. The disparity of all corresponding points in the scene will be adjusted by this amount. While the depth ordering of points is preserved through the translation, the horizontal translation shifts the perceived range of each scene point away from the viewer. The new perceived ranges can be obtained from (2) through the substitution of d_i with $d'_i = d_i - r$, where $r = d_{min}$ is the amount of the translation. The translation and cropping operation and the adjustment of the perceived ranges are depicted in Fig. 1b. If the perceived range of the closest point in the example scene had been behind the screen surface, we would perform the translation and cropping in the opposite direction and the perceived ranges would move closer to the viewer. Our task, therefore, is to estimate the disparity values within the given stereo-pair and to search for the disparity of the point closest to the viewer.

Since we are dealing with imagery to be presented to a human viewer, we assume the stereo-pair has been compressed to reduce storage and/or transmission requirements. Most common, interframe compression algorithms employ some form of block-based displacement estimation and compensation to improve coding performance.[1,2,3,4,13] This is the case for the recently proposed stereoscopic extension to the MPEG standard, which utilizes the temporal-scalability option of the standard to encode multiple views of the scene.[11,12] Block-based image prediction techniques assume that the differences between the two perspective views can be described by a piecewise, translational displacement; a single disparity value is assigned to a contiguous block of pixels. To utilize the information readily available from the compressed imagery, we base our eye-strain reduction algorithm on block-based disparity estimates that relate one view of the stereo-pair to the other.

The basic task of block-based disparity estimation is to find, for each block in the frame to be predicted, the relative location of the block in the reference frame that minimizes a given distortion function. For consistency, we will consider only the prediction of the right-eye image from the left-eye image. The minimum distortion block is searched for over a pre-specified horizontal search range (R). The search range often is selected heuristically based on assumptions of the maximum disparity in the stereo-pair. The disparity estimation process then can be described by,

$$\hat{d}(x_r, y_r) = \underset{\hat{d}(x_r, y_r) \in R}{\arg\min} \left\{ D\left[rblock(x_r, y_r), lblock\left(x_r - \hat{d}(x_r, y_r), y_r \right) \right] \right\} \tag{3}$$

where $rblock(\boldsymbol{p}_r)$ and $lblock(\boldsymbol{p}_l)$ are the right and left image blocks, $D[rblock, lblock]$ is the distortion between the two blocks, and $\hat{d}(x_r, y_r)$ is the estimated disparity value. Common distortion functions are the mean-squared error (MSE) and mean absolute difference (MAD) between the luminance components of the two blocks.

If only the actual disparity values within the image pair are obtained from the estimation process, the calculation of the disparity range would be trivial. However, estimation errors are likely to occur that will skew the range calculation. The accuracy of the disparity estimation process relies on numerous factors, including: 1) the block size, 2) the search range, 3) the degree of intensity uniformity of the pixels within block, and 4) the variation of the disparity throughout the image pair. Also, since the location of occluded regions is not known prior to the estimation procedure, meaningless disparity estimates will be calculated for these regions. For simplicity, we will group all erroneous disparity estimates as the result of occlusion. The actual disparity range then can be found by eliminating the contribution of all estimates obtained for occluded regions.

3. STATISTICAL MODEL AND RANGE CALCULATION

We take a probabilistic approach to the characterization of estimated disparity. Both the actual disparity and disparity estimates for occluded regions are modelled as random variables described by their probability density functions (pdf). The pdf of the estimated disparity is obtained through an application of Bayes' Law. A frequency interpretation estimate of this composite random variable's pdf can be obtained from the histogram of the estimated disparity values. From the histogram, we wish to obtain the pdf of the actual disparity values, which will yield the desired disparity range.

The distribution of the actual disparity within a stereo-pair is scene dependent. The probability of a specific disparity value is a function of both the relative size and the relative number of objects at a certain depth plane. Since the structure of the pdf of the actual disparity is difficult to describe, we avoid specifying an explicit pdf function and merely assume that it is a multimodal distribution.

The distribution of false disparity estimates is much easier to characterize. For occluded regions, the estimated disparity value that minimizes (3) will be equally likely for all locations within the given search range, and can be modelled as a uniformly distributed random variable. The probability of the estimated disparity given the estimate was obtained for an occluded regions then is given by,

$$Pr\{\hat{d}|\text{occluded}\} = \frac{1}{|R|}, \quad \hat{d} \in R \tag{4}$$

where, $|R| = x_{max} - x_{min} + 1$, is the inclusive length of the search range from x_{min} to x_{max}.

Letting $Pr\{\text{occluded}\} = \alpha = 1 - Pr\{\text{unoccluded}\}$, a direct application of Bayes' law yields the probabilities of the estimated disparity for both unoccluded and occluded regions over the search range as,

$$Pr\{\hat{d}\} = \frac{\alpha}{|R|} + (1-\alpha)\,Pr\{\hat{d}|\text{unoccluded}\} \tag{5}$$

While the quantities α and $Pr\{\hat{d}|\text{uoccluded}\}$ are unknown, we are not particularly concerned with these values. Instead, we are interested in finding the minimum value of $\hat{d}$ for which $Pr\{\hat{d}|\text{unoccluded}\}$ is non-zero. This can be accomplished through a simple thresholding of the histogram of the estimated disparity by $\frac{\alpha}{|R|}$. An occlusion detection algorithm could be used to estimate α from the stereo-pair's disparity vector map[7]; however, we feel this is an unnecessary and time-consuming step if we are merely interested in the disparity range. In our algorithm, we use the threshold,

$$\text{threshold} = \frac{K}{|R|}, \quad 0 < K < 1 \tag{6}$$

where an appropriate value for K is obtained through trial-and-error. Figure 2 shows the histogram of the disparity within a sample stereoscopic image. The dotted-line represents the threshold, where $K = 0.75$ The contribution from the erroneous estimates is effectively eliminated by the threshold and the actual disparity range of $[-15, 2]$ is accurately estimated. This diagram emphasizes the difficulty that would be experienced if our range calculation algorithm required an accurate description of the actual disparity's pdf.

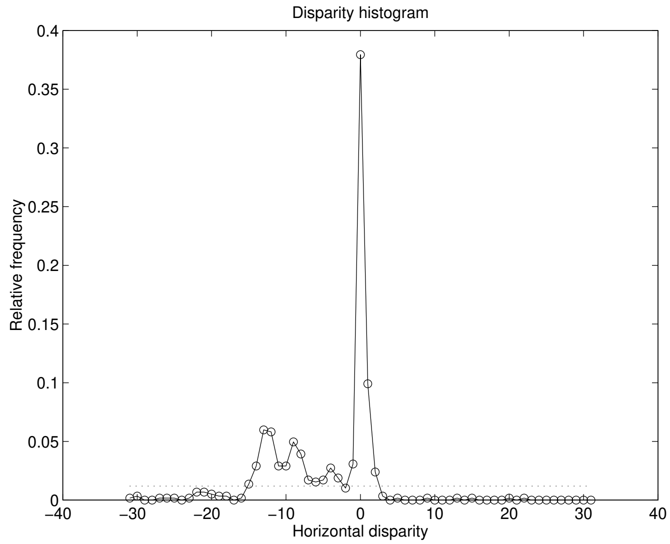


Figure 2: Sample estimated disparity histogram and threshold value.

If the eye-strain reduction algorithm is applied to a single stereo-pair, we simply threshold the estimated disparity histogram to find the most negative non-zero disparity probability and then translate the imagery by this amount. When we are provided with a sequence of stereo-pairs, we employ two enhancements to our basic technique.

The first enhancement attempts to eliminate high frequency noise in the two-dimensional histogram signal over time by applying a simple infinite-impulse response (IIR) low-pass filter to this signal.[9] The filtered histogram signal, $y(x, n)$, is given by,

$$y(x, n) = (1 - \beta_1) h(x, n) + \beta_1 y(x, n-1), \qquad x \in R \tag{7}$$

where $h(x, n)$ is the n^{th} frame's histogram, and β_1 is a damping factor ($0 < \beta_1 < 1$).

Since we speculate that abrupt and frequent transitions of the translation value will be visually objectionable, we would like to smooth the translation function over time. This is accomplished with the second enhancement, which is nearly identical to the first enhancement. The value of the image translation at frame n, $f(n)$, is a function of the previous frame's translation value and the current frame's minimum disparity value, $r(n)$, given by,

$$f(n) = (1 - \beta_2) r(n) + \beta_2 f(n-1) \tag{8}$$

where β_2 is another damping factor. We select both damping factors heuristically using the knowledge that as $\beta_i \to 0$ no filtering occurs.

4. EXPERIMENTAL RESULTS

The described eye-strain reduction algorithm was applied to a diverse set of real stereoscopic images and sequences. In this section, we present the quantitative results for two sequences. The sequences were compressed using a modified MPEG simulation capable of handling multiple views. The estimated disparity values were obtained from the compressed sequence bitstream. For all image pairs, block-based disparity estimation was performed using $16x16$ pixel blocks and the MAD distortion function. The search range for each sequence was selected to ensure that the actual disparity range of each image pair in the sequence was contained within the search range. Since a parallel camera configuration could not be ensured, the possibility

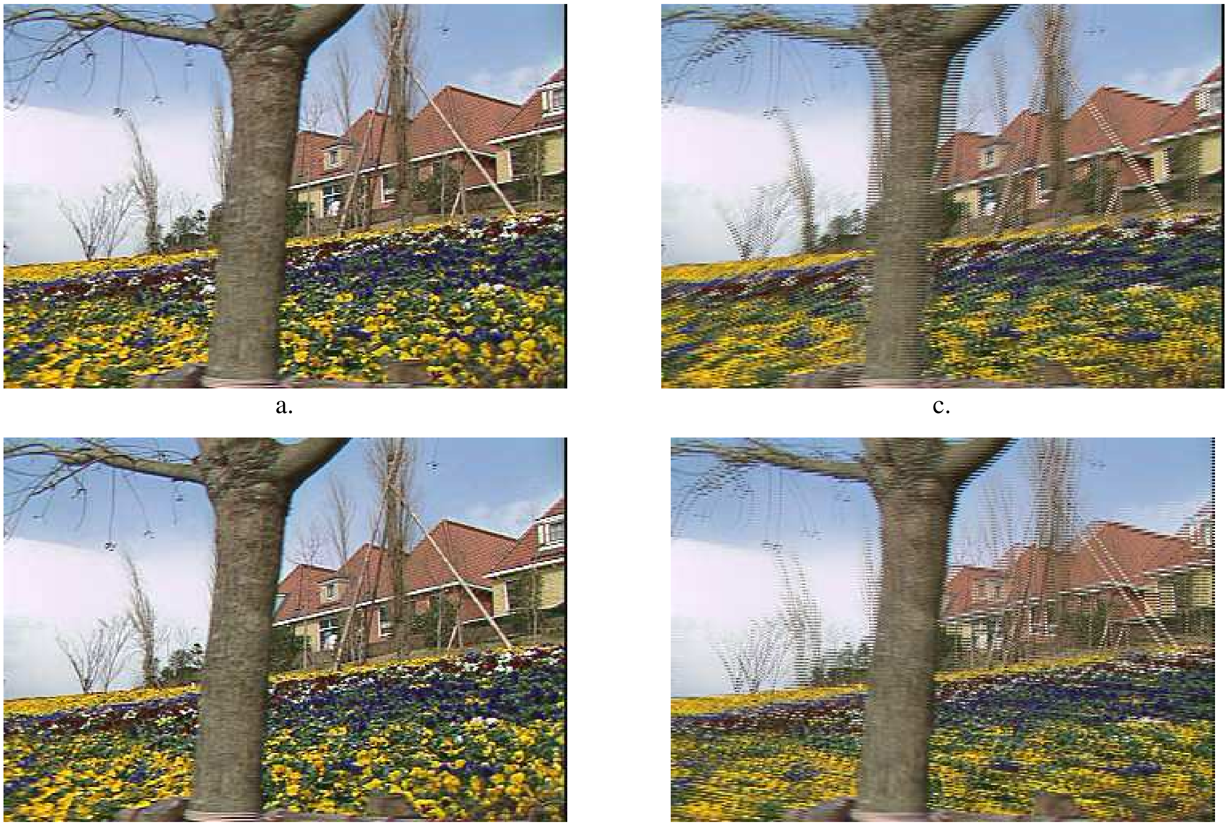


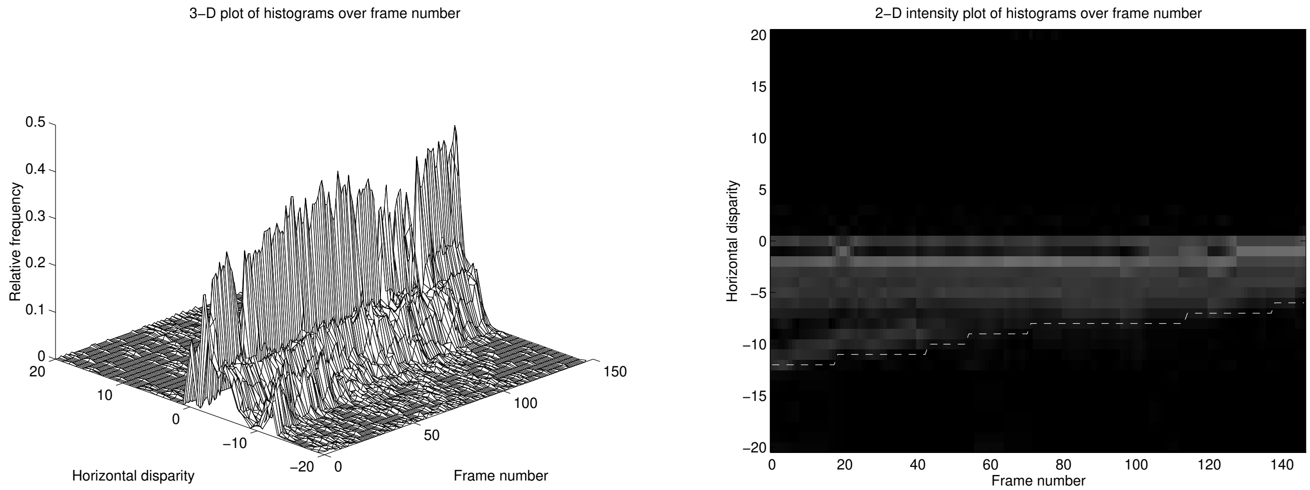
Figure 3: *Flower Garden* stereo-pair. a) original left-eye image, b) original right-eye image, c) unprocessed interlaced stereo-pair, d) interlaced stereo-pair after translation and cropping.

of minor vertical disparity was handled by extending the disparity search range to a two-dimensional grid. The horizontal disparity histograms were obtained by summing the 2-D histograms over the vertical disparity component. The parameters used in our algorithm were assigned the following values: $K = 0.75$, $\beta_1 = 0.5$, and $\beta_2 = 0.7$.

The performance of the algorithm in selecting the minimum disparity value was obtained by manually calculating the actual disparity range for various image pairs. The decoded bit-stream with and without image translation also was displayed on a stereoscopic-ready monitor. The subjective improvement of the eye-strain reduction algorithm was obtained through informal assessments by the authors and numerous visitors to our laboratory.

The first stereoscopic sequence examined, *Flower Garden*, was obtained from a monoscopic sequence with almost entirely horizontal camera motion. The left- and right-eye image sequences are the same monoscopic sequence delayed with respect to one another by two frames. The left- and right-eye images of the tenth stereo-pair of this sequence are shown in Figs. 3a and 3b, respectively. The line interlaced version of this stereo-pair, shown in Fig. 3c, illustrates the large negative disparity of the foreground tree. In fact, all image points in this stereo-pair have negative disparity, which make viewing of this sequence very difficult. A minimum disparity value of -12 pixels was calculated from the estimated disparity histogram for this image pair, and the resulting interlaced image after horizontal translation and cropping is shown in Fig. 3d. The foreground tree now appears to lie on the screen surface, with the rest of the scene behind the screen. While no formal human factors experiments were conducted, the raw and processed images were shown to approximately seven viewers, all of whom agreed that the processed imagery was much easier to view, and produced less eye-strain for them.

A three-dimensional plot of the individual stereo-pair histograms for the entire 146 frames of the *Flower Garden* sequence is shown in Fig. 4a. All of the image pairs are dominated by negative estimated disparity values, consistent with the visual

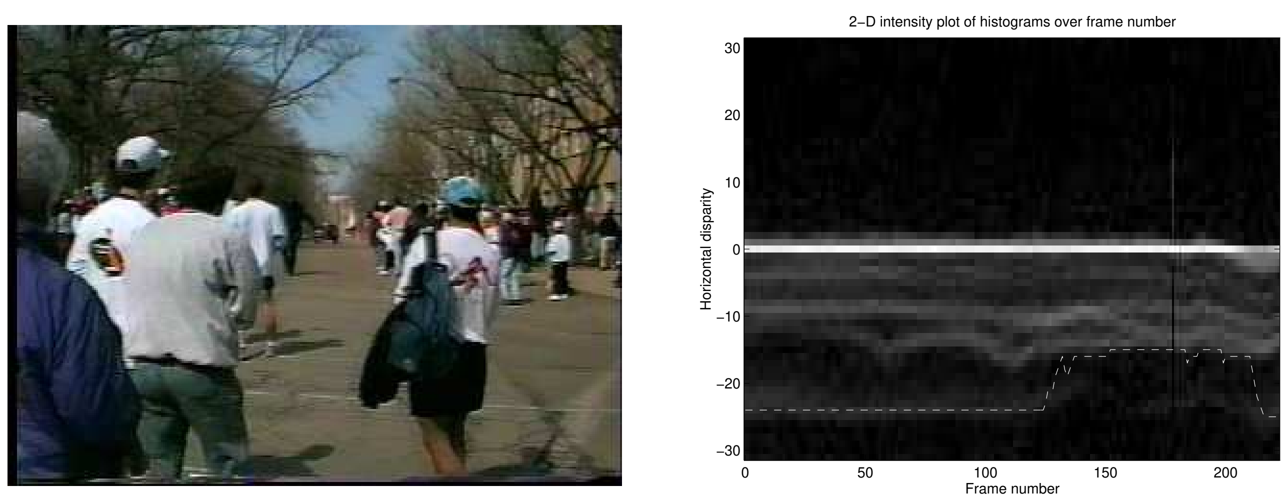


a. b.

Figure 4: *Flower Garden* disparity histograms. a) three-dimensional histogram plot, b) intensity plot of histograms and calculated translation values.

assessment that all objects appear to lie in front of the screen in the untranslated sequence. From this figure, one can observe the relatively small, non-zero disparity probabilities that occur outside of the major lobes. Without appropriate thresholding of these values, an inaccurate disparity range estimate would be obtained. This same information is presented in the form of a two-dimensional intensity image, in Fig. 4b, where white regions indicate relatively high disparity value probabilities. The dashed-line in this plot provides the horizontal translation value obtained from the eye-strain reduction algorithm. The upwardly sloping nature of the translation value is due to foreground objects, such as the large tree, moving out of the stereo-pair.

The second stereoscopic sequence examined, *Finish Line*, consisted of 223 stereo-pairs. The left-eye image of the 100^{th} pair is shown in Fig. 5a. Again, before the eye-strain reduction algorithm was applied, this sequence produced a great deal of eye-strain, confusion and depth ambiguity since all objects appeared to lie in front of the screen based on disparity. This was most troublesome for the object located on the left-side of the image, which is clearly clipped by the image surround. The intensity plot of the histograms for this sequence is shown in Fig. 5b. The histograms for the *Finish Line* sequence do not have as an abrupt edge as do the *Flower Garden* histograms. However, the minimum value was accurately estimated and eye-strain



a. b.

Figure 5: *Finish Line* sequence. a) original left-eye frame 100, b). histogram intensity plot and translation values.

was diminished. The positive and negative jumps in the translation value are the result of the person on the left-side of the image moving out of and back into the cameras' field of views. These transitions are relatively smooth and do not appear to be objectionable. While the histogram for frame 179 is completely dissimilar to its neighboring frames (due to a glitch in the original sequence), the filtering process of the algorithm maintains a constant translation value across this frame.

5. CONCLUSION

We have presented a simple algorithm for the reduction of the eye-strain often experienced when viewing stereoscopic imagery on a flat display device. Block-based disparity estimates relating the two views of the stereo-pair are used to form a histogram of the estimated disparity. This histogram is thresholded using a value derived from a statistical description of erroneous disparity estimates. The minimum disparity then is obtained from the modified histogram, and the image pair is translated to force all image points to lie on, or behind, the screen surface. Two simple enhancements of this algorithm, for stereoscopic sequences, also were presented; they rely on the assumption that the actual disparity range varies slowly over time.

The algorithm was shown to accurately select the minimum disparity value of real stereoscopic image pairs. Reduced viewer eye-strain and confusion were reported when the imagery was translated by the computed value. Formal human factors experiments are needed to verify these informal results. Even if future formal experiments invalidate our assumption that the most comfortable scenario is for all image points to lie behind the screen surface, the optimal image translation is almost certainly disparity dependent, so this algorithm will nevertheless be useful for accurately specifying the actual disparity range.

6. ACKNOWLEDGMENT

This work was supported by the Advanced Research Projects Agency under ARPA Grant. No. MDA 972-92-J-1010.

7. REFERENCES

1. ISO/IEC JTC1/SG29/WG11, ISO/IEC 11172-2, "Information Technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbits/s - Part 2: Video", May 1993.

2. ISO/IEC JTC1/SC29/WG11 Test Model Editing Committee, "MPEG-2 Video Test Model 5", ISO/IEC JTC1/SC29/WG11 Doc. N0400, April 1993.

3. ISO/IEC JTC1/SC29/WG11, "Information Technology - Generic Coding of Moving Pictures and Associated Audio, Recommendation H.262, ISO/IEC 13818-2, Draft International Standard," March 1994.

4. R. Chassaing, B. Choquet and D. Pele, "A stereoscopic television system (3D-TV) and compatible transmission on a MAC channel (3D-MAC)," *Signal Processing: Image Communication*, Vol. 4, No. 1, pp. 33-43, 1991.

5. V. S. Grinberg, G. W. Podnar and M. W. Siegel, "Geometry of binocular imagery," *Proc. SPIE Internat. Conf. on Stereoscopic Displays and Virtual Reality Systems*, Vol. 2177, pp. 56-65, San Jose, CA, 8-10 February 1994.

6. L. Lipton, *The CrystalEyes Handbook*, StereoGraphics Corporation, San Rafael, California, 1991.

7. J. S. McVeigh, M. W. Siegel and A. G. Jordan, "Adaptive reference frame selection for generalized video signal coding," *Proc. SPIE Internat. Conf. on Digital Video Compression: Algorithms and Technologies 1996*, Vol. 2668, pp. TBD, San Jose, CA, 31 January - 2 February1996.

8. T. Miyashita and T. Uchida, "Cause of fatigue and its improvement in stereoscopic displays," *Proc. of the SID*, Vol. 31, No. 3, pp. 249-254, 1990.

9. A. Oppenheim and R. W. Schafer, *Discrete-time signal processing*, chap. 5, Prentice Hall, Englewood Cliff, NJ, 1989.

10. S. Pastoor, "3D-television: A survey of recent research results on subjective requirements," *Signal Processing: Image Communication*, Vol. 4, No. 1, pp. 21-32, 1991.

11. A. Puri and B. Haskell, "Straw man proposal for multi-view profile," ISO/IEC JTC1/SC29/WG11 MPEG95/485, November 1995.

12. A. Puri, R. V. Kollarits and B. G. Haskell, "Stereoscopic video compression using temporal scalability," *Proc. SPIE Internat. Conf. on Visual Communications and Image Processing*, Vol. 2501, pp. 745-756, Taipei, Taiwan, 23-26 May 1995.

13. M. G. Perkins, "Data compression of stereopairs," *IEEE Trans. on Communications*, Vol. 40, No. 4, pp. 684-696, 1992.

14. C. Ware, C. Gobrecht and M. Paton, “Algorithm for dynamic disparity adjustment,” *Proc. SPIE Internat. Conf. on Stereoscopic Displays and Virtual Reality Systems II*, Vol. 2409, pp. 150-156, San Jose, CA, 7-9 February 1995.

15. T. Yamazaki, K. Kamijo and S. Fukuzumi, “Quantitative evaluation of visual fatigue encountered in viewing stereoscopic 3D displays: Near-point distance and visual evoked potential study,” *Proc. of the SID*, Vol. 31, No. 3, pp. 245-247, 1990.