SPECTRAL CONTINUITY AND EYE VERGENCE MOVEMENT

Lance R. Williams

Department of Computer Science
The Pennsylvania State University
University Park, Pennsylvania

ABSTRACT

In the Marr-Poggio model of human stereopsis, eye vergence movement is tightly coupled to the matching process. Any local area of any spatial frequency tuned channel can initiate a vergence movement designed to bring zero crossings within that local area into their range of correspondence. Since, in the human system, the resource of eye movement is limited, the control problem inherent in such a strategy seems intractable. This paper, in contrast, proposes that by requiring spectral continuity of disparity, global disambiguation can occur during the course of any reasonable sequence of vergence movements. The matching process is thus only loosely coupled to the vergence mechanism. A computer implementation has been tested successfully on both a random dot stereogram and a stereo pair of a natural scene.

1. Introduction

The structure of the physical world is constrained by physical laws, which in turn constrain images depicting it. The brain exploits these underlying constraints when it builds its world model from the images it receives from the eyes. Aii interesting conjecture is that it is able to do this with little or no *a priori* knowledge of semantic content. In this sense, early visual processing is a constructive process of description and is independent of recognition [1]. Perhaps the best example of the success of this approach is the Marr-Poggio model of human stereopsis, and its subsequent computer implementation by E.L. Crimson [2,3], Marr and Poggio propose that matching is conducted on zero crossings of the G' convolution of the left and right eye image. The results of matching zero crossings from low spatial frequency tuned channels are used to guide vergence movements which bring zero crossings from high spatial frequency tuned channels into their smaller range of correspondence. The vergence control mechanism is local, based on the success or failure of matching in local neighborhoods of each image.

The principal motivation behind this study is the desire to model the control of eye vergence movement in a biologically plausible manner. Crimson's implementation of the Marr-Poggio theory uses the results of matching within local areas of a channel to decide whether or not vergence movement is necessary to bring zero crossings within those local areas into correspondence. Local areas with less than 70% matches are declared out of range, and require vergence movement. The cross channel coupling is through the 21/2 -D sketch, which stores the results of matching from the low frequency channels. When a local area is out of range, these results are consulted to determine whether a convergent or divergent movement should be initiated.

However, there is an inherent difficulty with this control strategy, since vergence movement is a resource of limited access. Vergence movement is a physical process; the eyes are unable to make both convergent and divergent movements simultaneously. The single access resource of eye movement must however, satisfy the conflicting requirements of hundreds of different local areas within a single channel. The problem is further complicated because requests for eye movement can come from any local area of any channel, and must proceed sequentially from low frequency to high frequency.

This paper proposes a vergence strategy which is independent of the results of matching within local areas. The matching module is designed to take opportunistic advantage of any "reasonable" set of eye vergence movements. A reasonable set of eye vergence movements is defined to be any sequence of movements that spans the disparity range in a scene. In this implementation a single uniform vergence movement carries the eyes through a series of fixation positions, the optimal match over the range of the movement being preserved as in hysteresis. Vergence is thus only loosely coupled to the matching process, perhaps being controlled, as Marr and Poggio have suggested, by relative imbalances in the response of disparity sensitive pools [4], although on a global level. Kidd, Frisby and May hew's [5] demonstration that monocular cues can initiate vergence movement also supports the idea of the matching process being relatively independent of a specific vergence mechanism

If this simple vergence strategy is to prove sufficient, two objectives must be met; I) A consistent description of depth must be computed for a single vergence fixation; 2) This description must be integrated within the 21/2 -D sketch with other descriptions computed at other vergence fixations. Additionally, in the human system, the 21/2 -D sketch must be updated without explicit knowledge of eye position [2,3], which is probably not available. In order that these objectives may be examined properly, a review of the concept of *raw primal sketch* is in order.

2. Spectral Continuity of Disparity

Marr has suggested that one purpose of early visual processing is the creation of a symbolic description of physically meaningful changes in image intensity |6|. He developed a set of primal sketch tokens along with parsing rules based on cross channel combination of zero crossings. The first problem, that of creating the depth description for a single vergence fixation, can be solved by exploiting a binocular extension of Marr and Hildreth's |7] *spatial coincidence assumption:*

> If a zero-crossing segment is present in a set of independent V *G* channels over a contiguous range of sizes and the segment has the same position, orientation, and *measured disparity* in each channel, then the set of such zero-crossing segments may be taken to indicate the presence of an intensity change in the image that is due to a single physical phenomenon (adapted from [7])

Matching may occur in a more or less indiscriminate manner within each spatial frequency tuned channel, but only those matches which yield disparities of similar value to spatially coincident matches in neighboring channels are associated with the combined channel descriptor, or raw primal sketch token.

May hew and Erisby [8] have proposed that "the process of human binocular combination integrally relate the extraction of disparity information with the construction of raw primal sketch assertions." This work is therefore in great sympathy with their approach. However, it is worth noting the important difference

between this proposal and May hew and Frisby's proposal that "the patterns of between-channel correspondence., also help disambiguate within-channel fusions" Their program, FRECKLES, uses a cross channel description as an enriched matching primitive, allowing disambiguation of false targets to occur over a larger fusional area. The disambiguating power actually seems to stem from a form of *compatibility constraint;* matches are forbidden between cross-channel descriptions that could not have arisen from the same physical phenomenon [I].

In contrast, this paper proposes that a requirement of continuity of disparity in the frequency domain is used to select the optimal cross channel combination of matches obtained independently within each spatial frequency tuned channel. The intra-frequency matching is conducted in essentially the same manner as in the Marr-Poggio model, and like the Marr-Poggio model, the chief disambiguating power lies in the fact that the size of fusional area is restricted to the small interval in which false targets can be resolved by local neighborhood support [1,4]. Global disambiguation depends on the selection of the optimal cross channel combination during the course of vergence movement.

That there is an optimal cross channel combination of within channel matches follows from the fact that for each spatially localized area of the left or right eye image there will be a specific vergence fixation which will preserve spectral continuity optimally (Fig. 1). Since these matches are associated with a particular primal sketch token, the raw primal sketch provides a frame of reference for the maintenance of disparity values across vergence movements. The solution to the problem of updating the *21/2* -D sketch without explicit knowledge of eye position is thus implicit in the solution of the problem of correspondence in time between tokens of the primal sketch, a problem examined extensively by Ullman [9]

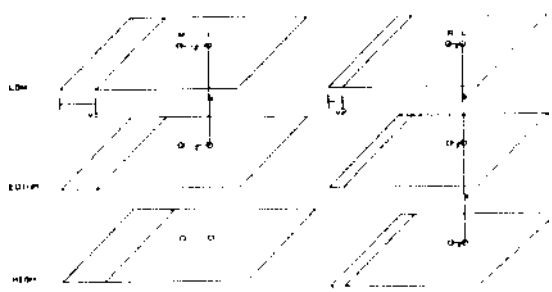

Figure 1. Vergence movement is simulated by increasing the displacement between-the left and right images. At the initial vergence position, v_1, only the low and middle frequency channels are in fusion. For this spatially localized area, optimal spectral continuity is obtained at v_2, when all three channels are in fusion. Thus, the matches at v_2 are associated with the combined channel descriptor, or primal sketch token.

This vergence strategy also has the advantage of eliminating the problem Crimson's program had solving images with periodic features [2,3]. The problem arose because the criterion for determining whether or not a vergence movement would be initiated was of a purely local nature. Any area of a channel with more than 70% matches would not initiate a vergence movement, irrespective of the results of matching within spatially coincident areas of lower frequency channels. Thus, different initial vergence positions could produce different matching results. By requiring that disparity remain relatively constant for a given feature through the frequency domain, the more local matches of higher frequency channels are forced to agree with the more global definition of disparity provided by lower frequency channels. As long as the lowest frequency channel is of lower spatial frequency than the period of the repeating pattern (windows on an office building in the case of Crimson's program) correct disparity values will be assigned.

3. Implementation

The program proceeds by first extracting zero crossings for both eyes at each spatial frequency. This is accomplished by traversing each horizontal raster looking for sign changes in the $\nabla^2 G$ profile. The current implementation uses two spatial frequency tuned channels. Each channel is processed independently, the low frequency channel first (in the human system, intra-frequency matching would be conducted concurrently within all channels). Vergence movement is simulated by incrementally increasing the displacement between the left and right images by an amount equal to half the fusional area of the high frequency channel (the actual amount is not critical). This has the effect of moving the plane of fixation through the disparity range determined by the matches of the low frequency channel. Matching is repeated at each fixation, only matches yielding disparities consistent with those of the lower frequency being accepted. For the two channel implementation, minimal spectral continuity is also optimal spectral continuity.

Matching proceeds within a frequency by first forming "competition" matrices from the cross product of like signed zero crossings from corresponding rasters of the left and right images (Fig. 2). Zero crossings from the right eye form rows while zero crossings from the left form columns. All potential matches are explicitly represented by a unique position in the competition matrix and associated with each row and column is the x-eoordinate corresponding to the location of the zero crossing along the raster of the appropriate image. A disparity is assigned to each potential match by subtracting the x-coordmates of all rows and columns. Those potential matches having disparity smaller than *2a* for the channel being processed are marked as targets.

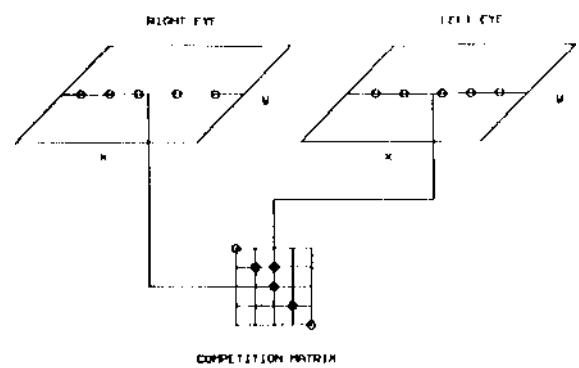

Figure 2. A competition matrix is formed from the cross product of like signed zero crossings from corresponding rasters of the left and right image. Potential matches with disparity less than *'la* are marked with circles. A match is ambiguous when more than one match appears in any given row or column. Ambiguous matches are resolved by the pulling effect, which favors matches of similiar disparity to surrounding ambiguous matches.

Targets are ambiguous when more than one target is present in any given row or column. Two targets in the same column represent ambiguity to the left eye, while two targets in the same row represent ambiguity to the right. Every raster is first processed for unambiguous matches and the results stored in a buffer so that the *pulling effect,* as in the Marr-Poggio model, may use these results to resolve ambiguous targets. Before a match from a high frequency channel is accepted, the majority disparity within a cir-

cular region surrounding the spatially coincident area of the low frequency channel is calculated. Only those matches having disparity within a small c (equal to *a* for the channel) of the majority disparity are accepted as correct. The entire matching process is then repeated at the next vergence fixation.

4. Conclusion

A computer implementation has been tested on a random dot stereogram (Fig. 3,4) and a stereo pair of a natural scene (Fig. 5,6) with good qualitative results. The vergence strategy was very simple and consisted of moving the plane of fixation through the entire disparity range in a single uniform movement. Global disambiguation was effected by requiring that matches from spatially coincedent areas of adjacent spatial frequencies possess similar disparity. This constraint is a binocular extension of the spectral continuity property of raw primal sketch tokens. Thus it has been shown that there is no need for local control of the vergence mechanism and its associated problems. Additionally, it has been suggested that the raw primal sketch provides a frame of reference for the maintenance of disparity values across vergence movements.

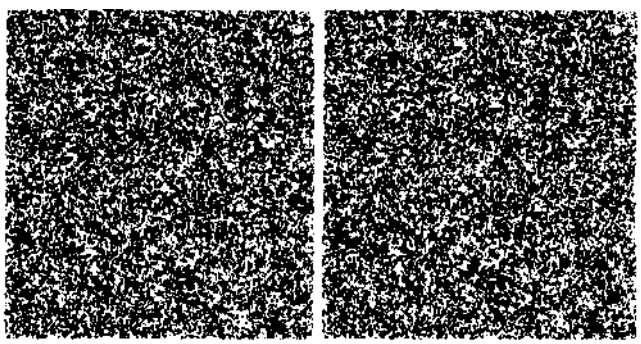

Figure 3. The random dot stereogram used in this study.

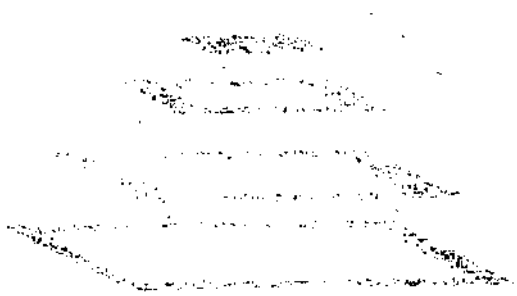

Figure 4. The solution for the random dot stereogram.

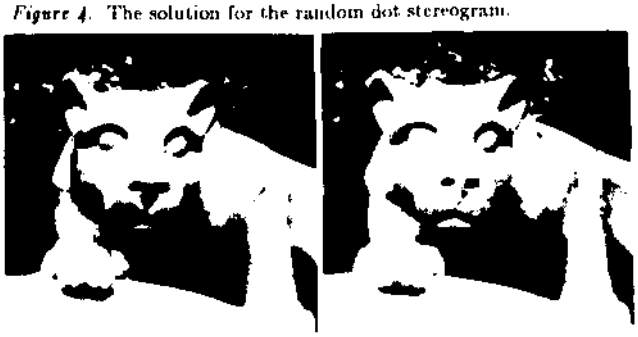

Figure 5. Stereo pair of the Nittany Lion shrine. Each image is 256x256 with 16 grey levels.

Figure 6. Disparity map for the Nittany Lion image.

ACKNOWLEDGMENTS

Special thanks to Bill Sakoda. Thanks also to Tony Maida, Gordon Shulman, John Sacha, and Jon Crouse. This work was supported in part through a grant from the General Electric Corporation.

REFERENCES

[I] Marr, D. 1982. *Vision.* San Francisco. W.H. Freeman and Company.

|2| Crimson, VV.E.L. 1980. A computer implementation of a theory of human stereo vision. MIT A.I. Lab. Memo 505. "Phil. Trans. Soc. Lond. B292" ,217-253.

|3| Crimson, W.E.L. 1981. *From Images to Surfaces.* Cambridge, Mass.: MIT Press.

j4] Marr, D., and T. Poggio. 1979. A computational theory of human stereo vision. "Proc. R. Soc. Lond. B204",301-328.

[5| Kidd, A.L., J.P. Frisby, and J.E.W. Mayhew. 1979. Texture contours can facilitate stereopsis by initiating appropriate vergence eye movements, "Nature 280",829-832

[Q Marr, I). 1970. Early processing of visual information. "Phil. Trans. R. Soc. Lond. B275" ,483-524.

[7] Marr, D., and E. Hildreth. 1980. Theory of edge detection "Proc. R. Soc. Lond. B207",187-217.

|8| Mayhew, J.E.W., and J.P. Frisby. 1981. Psychophysical and computational studies toward a theory of human stereopsis. "Artificial Intelligence 17", 349-385.

[9] Ullman, S. 1979. *The Interpretation of Visual Motion.* Cambridge, Mass.: MIT Press.