

STEREO RECONSTRUCTION THROUGH DISPARITY SPACE

Richard Baldwin
 Department of Land Surveying
 Polytechnic of East London
 Longbridge Road, Dagenham
 Essex RM8 2AS UK

Katsuhiko Sakaue, Hiromitsu Yamada and Kazuhiko Yamamoto
 Image Understanding Section
 ElectroTechnical Laboratory
 Tsukuba Science City
 Ibaraki 305 Japan

email baldwin%etl.go.jp@RELAY.CS.NET

ABSTRACT

This paper is concerned with surface reconstruction from stereo image data. We do not explicitly match left and right image points, but use a new formulation of *Disparity Space* geometry to recover the object surface in terms of very dense *Intensity* and *Disparity maps*. We introduce Disparity Space, and develop an approach using multiple matching primitives to form *Local cost functions* for which a *Global solution* is achieved using *Dynamic Programming*. Disparity resolution is controlled by scale factors, and the algorithm can be implemented in parallel. Experimental results with close range and underwater images are reported.

1. INTRODUCTION

Passive stereo has an almost instantaneous data capture stage which can usefully store large amounts of 3D spatial and scene descriptive information. This information can be recovered through an *image inversion* operation. Image formation and inversion involve both photometric and geometric processes. The geometric processes (lens distortion, camera position and orientation, relief displacement) are well understood, and result in an object point being projected to a particular image point, according to a known camera geometry model. The effects of the photometric processes (illumination, reflectance) are more complicated as they are dependent on geometry and local photometric factors. The photometric model has not been considered in the present work.

We seek to recover the spatial information through the calculation of very dense range and intensity maps using digital stereo image data. We can represent this *image inversion* procedure as:

$$L_{(x_1, y_1, i_1)}, R_{(x_2, y_2, i_2)} \Rightarrow \begin{cases} I_{(X, Y)} \\ D_{(X, Y)} \end{cases} \quad (1)$$

that is, given an image pair $L_{(x_1, y_1, i_1)}, R_{(x_2, y_2, i_2)}$ of known camera geometry, we recover the surface as a *Disparity map* $D_{(X, Y)}$ and an *Intensity map* $I_{(X, Y)}$. We have developed a new approach which applies *global constraints* to *local cost functions* calculated using a number of different matching primitives for a range of disparity values. The global solution is obtained by applying geometric constraints and Dynamic Programming (DP) in Disparity Space which yields a disparity value for every DP node.

2. DISPARITY SPACE

2.1 Introduction and theoretical background

Consider a stereo setup as shown in Figure 1a with cameras of identical geometric and electro-optic behaviour mounted with their optical axes exactly parallel with separation b and principal distance f . We introduce parallel image coordinate systems (x_1, y_1) and (x_2, y_2) for the left and right images respectively, and a 3D coordinate system (XYZ) centred upon a point midway between the two perspective centres (similar to Horn 1986). As the images are parallel, the epipolar constraint means that a point located at (x_1, y_1) must have a conjugate point located at (x_2, y_2) where y_1 and y_2 are equal and x_1, x_2 differ by an amount referred to as the disparity of the point.¹

¹any stereopair can be transformed to parallel images, given the camera internal geometry, and a minimum of five well distributed point correspondences

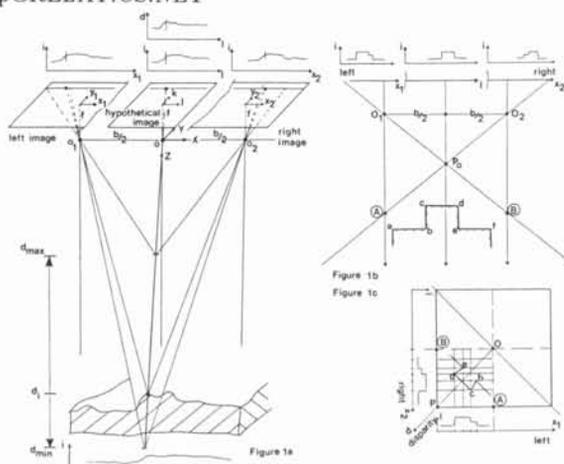


Figure 1: Introduction of Disparity Space. All points located on the central image l, k have an associated disparity value d and an intensity value. We specify a coordinate (l, k) and then for all values of d within the disparity range, we calculate the matching function specified at (x_1, x_2) . The results are stored in disparity space.



Figure 2: Disparity Space showing $(i_1 - i_2)^2$. Using a single, simple matching function, there are many potential matches.

The introduction of Disparity Space can be physically considered as the introduction of a hypothetical image located midway between the left and right images. We introduce such an image (Figure 1a) and define its image coordinate system as (l, k) which is parallel to $(x_1, y_1), (x_2, y_2)$ and the (XYZ) system. This hypothetical image consists of two separate images; an intensity image $I_{(l, k)}$ and a disparity image $D_{(l, k)}$. From Figure 1a, we have:

$$(x_2 - x_1) = bf/Z \quad (2)$$

The images are parallel and so conjugate image points $(x_1, y_1), (x_2, y_2)$ will form an image point in the hypothetical image at (l, k) with disparity defined as d where:

$$\begin{aligned} l &= \lambda_1(x_2 + x_1)/2 \\ k &= \lambda_1(y_2 + y_1)/2 \\ d &= \lambda_2(x_2 - x_1)/2 \\ f_c &= \lambda_1 \cdot f \end{aligned} \quad (3)$$

λ_1, λ_2 are scale factors and f_c is the effective focal length. Expressing

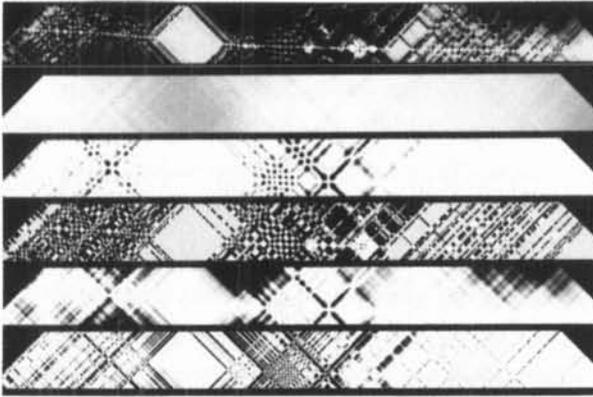


Figure 3: Disparity Space Stereo correspondence. a) sum of local costs. b) average intensity. c) LOG magnitude match. d) edge orientation match e) intensity match. f) edge magnitude match

the first and third of eqs (3) in matrix form, we have:

$$\begin{bmatrix} l \\ d \end{bmatrix} = 1/2 \begin{bmatrix} \lambda_1 & \lambda_1 \\ -\lambda_2 & \lambda_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (4)$$

The inverse form is:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1/\lambda_1 & -1/\lambda_2 \\ 1/\lambda_1 & 1/\lambda_2 \end{bmatrix} \begin{bmatrix} l \\ d \end{bmatrix} \quad (5)$$

Further, $l = f_c \cdot X/Z$, $k = f_c \cdot Y/Z$ and using eqs (2) & eqs (3):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{b \cdot \lambda_2}{2d \cdot \lambda_1} \begin{bmatrix} l \\ k \\ f_c \end{bmatrix} \quad (6)$$

Equations (4) and (5) represent a two dimensional affine transformation of the search plane defined by x_1 and x_2 . The new (l, d) system we call *Disparity Space* and we note there is a corresponding value of (l, d) for every value of (x_1, x_2) . The values of the λ_1, λ_2 scale factors control the resolution of the disparity map and the separation of the evaluation nodes.

2.2 Disparity Space stereo correspondence

A continuous object surface must form a continuous surface in Disparity Space (Fig 1c). Each point (l, k) defines a position upon the central image and hence a vector in space. This vector must cut the object surface as defined by eqs (6) and so the problem reduces to that of determining d for all points (l, k) . Further, it is apparent that we can place limits on the possible values of d , that is $d_{min} \leq d \leq d_{max}$ for all possible matches, and it is unnecessary to insist that the solution takes place along an epipolar line. We can specify any logical succession of (l, k) , place limits on d and evaluate all possible local matches. In this paper we have performed evaluation along and at right angles to the epipolar line.

Disparity Space stereo reconstruction takes place as follows. For all points (l, k) , *local cost functions* are evaluated for every legal value of disparity. This local matching will not always yield an unambiguous result (Fig 2), hence we need a global constraint in order to resolve the local ambiguities. This is especially true for images with slow intensity changes, poor surface texture or highlight. We need to consider the choice of local matching primitives and also how to achieve a global solution.

3. LOCAL COST FUNCTIONS

3.1 Choice of Matching Primitive

The choice of matching primitives has received much attention in the literature and distinction is usually drawn between methods using *feature based* primitives (Marr & Poggio, 1979, Hildreth, 1983) and

area based primitives (Ackermann, 1984; Gruen, 1985) which is also reflected in the matching strategy. Area based correlation methods can produce a dense, highly accurate DEM, but require good initial values, have a small pull-in range and are sensitive to textures and slow intensity changes. Feature based methods rely on the extraction of tokens, (eg edges) and a constraint satisfaction algorithm. There may be some assumption relating the extracted token to the object surface; problems with correctly detecting only significant features, tokenisation and surface interpolation. The method can quickly yield sparse depth information.

It is apparent that certain image types respond more favourably to the use of certain matching primitives. Hence we suggest that a general purpose stereo system should allow flexibility in the choice of the local matching primitive.

3.2 Combining matching primitives

We consider every pixel to possess a set of matching attributes. These attributes can be flexibly defined, provided that they can be considered to act at a pixel; are considered as observed quantities and the expectation of the matching function is zero in the event of a successful match. The output of linear filters conforms to these conditions². The output of any area based correlation function can be made to comply with these requirements, assuming the function is considered to act at an evaluation node. We linearly combine different matching functions together. Our implementation uses

$$local.match.cost = \Lambda \cdot [(I_1 - I_2)^2/\sigma_1 + (\partial I_1 - \partial I_2)^2/\sigma_2 + (\partial^2 I_1 - \partial^2 I_2)^2/\sigma_3 + (\Theta_1 - \Theta_2)^2/\sigma_4] \quad (7)$$

$$\Lambda = \begin{cases} 1 & \text{for no constraint} \\ \sqrt{\frac{d_{(l,k)} - d_{(l,k-1)}}{d_{(l,k)} - d_{(l-1,k)}}} & \text{for epipolar solution} \\ \sqrt{\frac{d_{(l,k)} - d_{(l,k-1)}}{d_{(l,k)} - d_{(l-1,k)}}} & \text{for solution at right angles to epipole} \end{cases} \quad (8)$$

where Λ is an additional weight defined as the departure from the adjacent solution node and $\sigma_1, \dots, \sigma_4$ are function weights derived from the variance of the individual matching functions. Figure 3 shows a section through Disparity Space calculated using eq (7). In many stereo matching systems a set of disambiguating functions are introduced to prevent false matches. In our system, using such a rich variety of input data we do not need such disambiguating functions. Any local ambiguity is resolved by the use of multiple matching primitives and global constraints.

4. DYNAMIC PROGRAMMING (DP)

4.1 DP for Global Solution & previous work

Dynamic Programming is a method for solving nonlinear optimisation problems (Bellman, 1957), where the solution can be expressed as a combinatorial sequence of decisions based upon a small number of potential solutions at each evaluation node. Implicit within the algorithm is the concept of *ordering* whereby a solution proceeds smoothly through a sequence of ordered subproblems within the finite search space. Each subproblem has only a small number of possible solutions, the solution to which is based upon comparing *local cost factors* associated with each possible solution to the subproblem. Each subproblem is solved in term, by choosing the optimum (ie least cost) solution. Possible solutions are defined in terms of *transition rules* which specify the set of possible decisions (or solutions) valid at each subproblem stage. A record is kept of the solution selected for each subproblem. Once the decision mechanism has traversed the search space, then the solution record is examined and the global solution is constructed by backtracking through the decisions of each evaluation stage.

Bernard (1984), Ohta & Kanade (1985), Lloyd (1986), Lloyd, *et al* (1987), Kolbl & da Silva (1988) have all used Dynamic Programming in a 2d (x_1, x_2) search space to solve globally by sections defined along epipolar lines. Additional constraints between epipolar lines were proposed by Ohta & Kanade (1985) and Lloyd (1986).

²Note that there may be some distortion of the filter output (Nishihara, 1983)

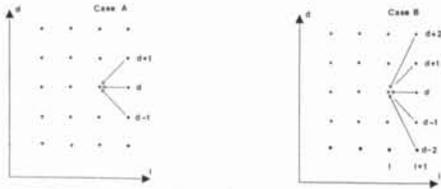


Figure 4: DP transition rules. a) Case A - simplest case for eqs (9). b) Case B - modified transition rules (eg for $\lambda_1 = 1, \lambda_2 = 2$)

4.2 DP in Disparity Space

The relationship between the 2D search space (l, d) and the $(x_1, y_1), (x_2, y_2)$ images is described by eqs (4) and (5). We can define any succession of l, k and then evaluate $(x_1, y_1), (x_2, y_2)$ for a given range $d_{min} \leq d \leq d_{max}$. The local match cost is calculated according to eq (7) and stored (Fig 3). The global DP solution is determined by calculating the cumulative cost according to a set of transition rules and a suitable recurrence relationship.

We have experimented with several sets of transition rules and recurrence relationships, defined to act along epipolar lines and also at right angles to these lines. Eqs (9) show one example of the recurrence relationship defined for the simplest epipolar solution. The associated transition rules are shown in Case A of Fig 4a.

$$cumcost_{l,d} = Min \begin{cases} cumcost_{l+1,d+1} + \beta_1 \cdot localcost_{l,d} \\ cumcost_{l+1,d} + \beta_2 \cdot localcost_{l,d} \\ cumcost_{l+1,d-1} + \beta_1 \cdot localcost_{l,d} \end{cases} \quad (9)$$

5. EXPERIMENTS AND RESULTS

Our approach has been tested with close range and aerial images (Baldwin & Yamamoto 1990). Owing to lack of space, only the close range results are reported here. Each image is transformed into an epipolar image (see Section 2.1), smoothed using a Gaussian filter and smoothed intensity gradient magnitude and orientation is obtained using a Sobel operator. Direction is quantised as 8 bit data. Finally, the original image is operated upon by a Laplacian of Gaussian (LOG) to produce the second derivative. Various filter sizes were tried for the LOG filter. All computations were performed on a Mips R3000 RISC machine with R3010 fpu.

Venus Images

These are the first images that we have used the algorithm with. Here we were interested in how the algorithm would cope with very slow intensity changes, smooth, almost featureless surfaces and high-light. As it is the stereomatching that we were interested in, no camera calibration was carried out. The disparity space solution at right angles to epipolar lines is shown in progress in Figure 5. where the high quality of the reconstructed central image $I_{(l,k)}$ is apparent. Note that the computed central image can be fused with both the left and right original image. Total computation time for the 256 x 256 image varies from 3 minutes for $\lambda_1 = 1, \lambda_2 = 1$ to about an hour for $\lambda_1 = 4, \lambda_2 = 2$ giving an effective depth resolution of two pixels and 0.25 pixels. See Fig 6 for results with various values of λ .

Underwater Images

Enlarged prints were made from an underwater stereopair taken with a Photosea 2000 35mm stereocamera ($b = 250$ and $f = 49$ mm approx), digitised using a Sharp color scanner at 150 dots per inch, and then 680 x 680 pixel images (with 8 bit gray scale) were extracted. This is equivalent to scanning the original 24 x 36 mm negative at approximately 750 dots per inch. This image pair has an abundance of fine texture, though contains highlight, shadow, significant lens distortion, perspective distortion, areas of poor definition, a large disparity range and forms an incomplete model. 3D views of the reconstructed surface are shown in Figure 7. Total computation time for the 462,400 points varies from 68 minutes for $\lambda_1 = 1, \lambda_2 = 1$ to approximately ten hours for $\lambda_1 = 2, \lambda_2 = 2$ giving an effective depth resolution of two pixels, and 0.5 pixels. Note that by changing the projection and using surface rendering, the results can be presented at any suitable orientation.

6. DISCUSSION

This paper has introduced a new stereo matching technique which is flexible in its use of matching primitives and presents a unified framework for the recovery of dense disparity and intensity maps. There are a number of interesting aspects to this work.

Firstly, we have been able to present the framework of Disparity Space, and show how it can be associated with a hypothetical middle image Intensity map $I_{(l,k)}$, which we have then recovered as an integral part of the solution. In some respects, the current work offers a unification of the global disparity field work of Witkin, *et al* (1987) and Barnard (1989), and the epipolar correspondence of Ohta & Kanade (1985) and similar work. The *Disparity Gradient* limit of Pollard, Mayhew & Frisbee (1985) is also included, as this simply represents a transition rule.

Secondly, the idea of trading scale for depth precision (Nishihara, 1983) is incorporated through the use of the scale factors λ_1 , and λ_2 , and we have presented some results at different scales. Note that the increased precision is achieved at the expense of a very significant increase in computation time when using a serial machine. The algorithm can be recast as a coarse-fine parallel algorithm, by iteratively changing the λ values and then solving for each solution line independently and concurrently.

Thirdly, the properties of Disparity Space itself are interesting, though not explored in this paper. We note (Fig 1c) that there exists a point in Disparity Space for all points in the stereo envelope. The search space is finite, and represents all points in the object space from infinity O to the closest point of the stereo overlap P. All vertical lines pass through the point O, and two points located at the same distance form a line at right angles to PO. Thus we see that there are spatial relationships implicit within Disparity Space geometry. It may be interesting to develop other 'Shape from X' algorithms in terms of Disparity Space, as it may offer a framework for unifying disparity and other depth and shape cues.

The Collinearity equations can be derived by introducing $Xg = s.R.X + \Delta X$ where Xg is the ground system; R is a rotation matrix; s is a scale factor and substituting for the model coordinates X using eq(6). ΔX are shift terms. The above equation can be used to determine ground coordinates in any arbitrary system, given a minimum of three control points. Hence the Disparity map $D_{(l,k)}$ and the Intensity map $I_{(l,k)}$ can be transformed to a ground value and so we have solved eq (1).

Finally, control points, straight lines, planar surfaces can be projected into Disparity Space and then used to aid the global solution. We should reconsider the nature of control in this kind of work.

7. CONCLUSIONS

We have presented a new method for stereo reconstruction that combines various matching functions within the representational framework of Disparity Space that can be regarded as a unification of earlier work. The system can be solved at different scales, and we have used a global Dynamic Programming solution to solve for sections through the defined space system, which can be specified as any logical route through disparity space. The method has been applied to close range and aerial images which exhibit a variety of image problems, such as poor surface texture, smooth intensity changes, repetitive texture, highlight, shadow and incomplete models.

In future, we wish to include a photometric model and control features within the solution; we also wish to perform a detailed accuracy evaluation; and the non-epipolar solution is expected to have some success with occlusions.

ACKNOWLEDGEMENTS

This work was carried out through the award of a Japanese Science and Technology Agency Fellowship tenable in the Image Understanding Section at ETL. Thanks are due to S. Muraki and G. Wolberg (Columbia University) for the rendering routine.

REFERENCES

- ACKERMANN, F., 1984. Digital image correlation: performance and potential applications. *The Photogrammetric Record*, 11(64):429-439.
- BALDWIN, R.A. and YAMAMOTO, K., 1990. Digital Superimposition and automatic DEM production by Dynamic Programming in Disparity Space. *ISPRS*, vol 28(4):259-272.
- BARNARD, S.T., 1989. Stochastic Stereo Matching over Scale. *IJCV* (3):17-32.
- BELLMAN, 1957. *Dynamic Programming*. Princeton University Press. USA.
- BERNARD, 1984. Automatic stereophotogrammetry: A method based on Feature detection and DP. *Photogrammetria* (39):169-181.
- GRUEN A., 1985. Adaptive LS correlation: A powerful image matching technique. *S. African J. P & R.S.* 14(3):175-87.
- HILDRETH, E.C., 1983. The detection of intensity changes by computer and biological vision systems. *CVGIP* (22):1-27.
- HORN, B.K.P., 1986. *Robot Vision*. The MIT Press, Macgraw Hill. New York.
- LLOYD, S.A., 1986. Stereomatching using intra- and inter-row Dynamic Programming. *Pattern Rec. Letters* (4):273-277.
- LLOYD, S.A., HADDOW E.R. and BOYCE, J.F., 1987. A parallel binocular stereo algorithm using DP & Relaxation labelling. *CVGIP* (39):202-225.
- KOLBL, O. and Da Silva, I., 1988. Derivation of a digital terrain model by Dynamic Programming. *ISPRS* 27(B8):III 77-86.
- MARR, D. 1982. *Vision*. W.H.Freeman. San Francisco. USA. 397 pages.
- MARR, D., and POGGIO, T. 1979. A theory of human stereo vision. *Proc. Royal Soc.London* (204):301-328.
- MAYHEW J.W. and FRISBY, J.P. 1981. Psychophysical and computational studies towards a theory of human stereopsis. *AI* (17):349-385.
- NISHIHARA, K., 1983. PRISM: A practical realtime imaging stereomatcher. *Proc SPIE* 449:134-142.
- OHTA, Y. and KANADE, T., 1985. Stereo by intra- and inter- linescan search using Dynamic Programming. *IEEE Trans. PAMI*, 7(2): 139-154.
- POLLARD, S.B., MAYHEW, J.W., and FRISBEE, J.P., 1985. PMF: a stereo algorithm using a disparity gradient limit. *Perception* (14): 449-470.
- WITKIN, A., TERZOPOULOS, D., KASS, M., 1987. Signal matching through scale space. *IJCV* (1):133-144.

Figure 5: Venus Solution in progress for $\lambda_1 = 2, \lambda_2 = 2$ using transition rules A. The reconstructed Intensity map $I_{(l,k)}$ and the Disparity map $D_{(l,k)}$ are shown almost completed. Note that both the left and right image will fuse with the center image for stereovision

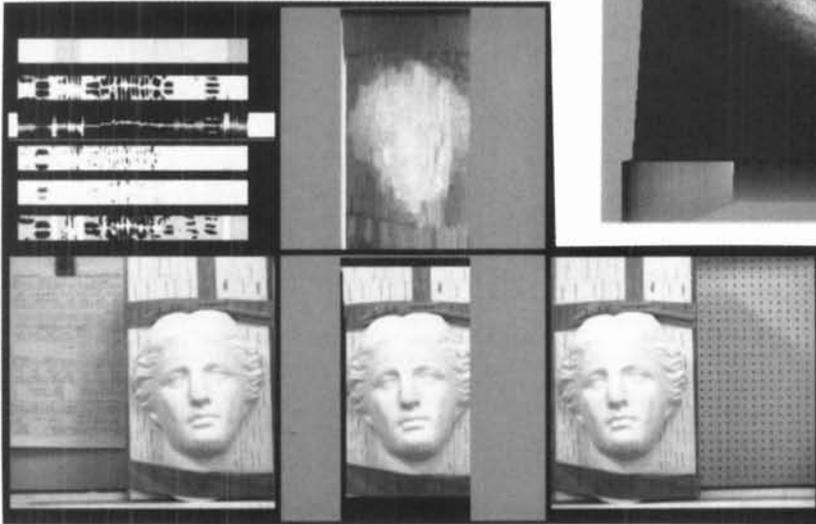


Figure6: Venus solutions for different depth resolution.

a) $\lambda_1 = 1, \lambda_2 = 1$

b) $\lambda_1 = 1, \lambda_2 = 2$

c) $\lambda_1 = 2, \lambda_2 = 2$

d) $\lambda_1 = 2, \lambda_2 = 4$

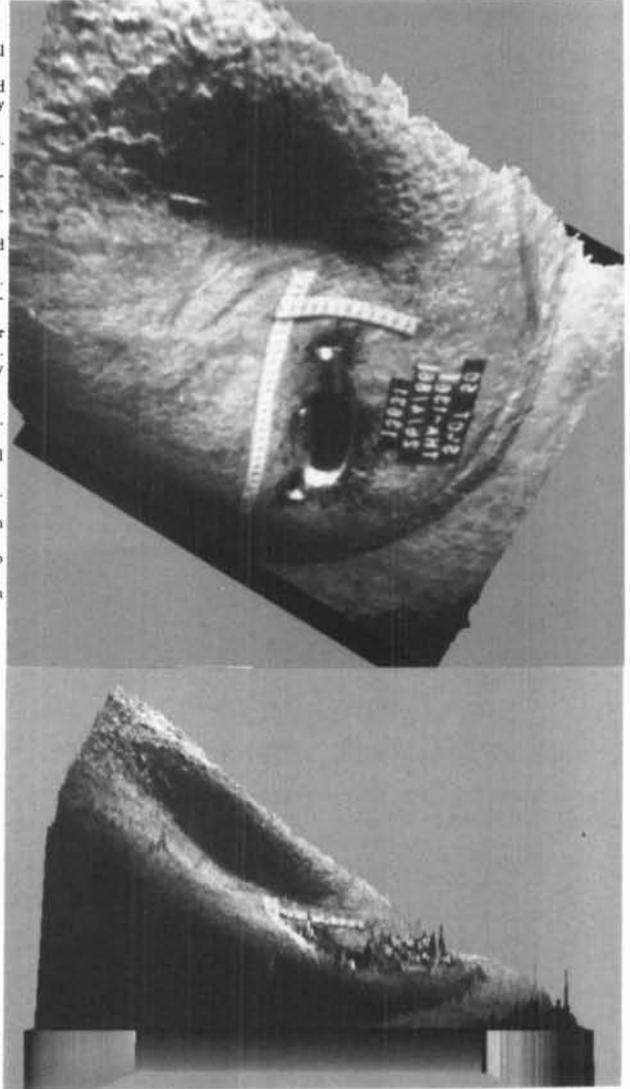


Figure 7: Surface reconstruction of damaged underwater pipe. The Intensity map $I_{(l,k)}$ and the Disparity map $D_{(l,k)}$ have been calculated by the stereo reconstruction method presented in this paper, and then the above perspective views have been created by rendering the surface with the Intensity map. The above are raw results, there has been no editing, smoothing or point rejection. The lower image clearly shows the depth errors concentrated in the area of the puncture through the pipe and the area around the text.

