Proceedings of DARPA Image Understanding Workshop, Jan. 27-29, 1992
pages 409-426

A Multiple-baseline Stereo Method *

Takeo Kanade **Masatoshi Okutomi†** **Tomoharu Nakahara**
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA

Part I: Theory

M. Okutomi and T. Kanade

Abstract

This paper presents a stereo matching method which uses multiple stereo pairs with various baselines to obtain precise distance estimates without suffering from ambiguity.

In stereo ptvcessing. a short baseline means that the estimated distance will be less precise due to narrow triangulation. For more precise distance estimation. a longer baseline is desired With a longer baseline. however. a larger disparity range must be searched rofind a march. As a result, marching is more difficult and there is a greater possibility of a false match. So there is a trade-off between precision and accuracy in matching.

The stereo matching method presented in this paper uses multiple stereo pairs with different baselines generated by a lateral displacement of a camera Matching is performed simply bv computing the sum of squared-difference (SSD) values. The SSD functions for individual stereo pairs are represented with respect to the inverse distance (rather than the disparity as is usually done). and then are simply added to produce rhe sum of SSDs. This resulting function is called the SSSD-in-inverse-distance. We show that the SSSD-in-inverse-distance function exhibits a unique and clear minimum at the correct matching position even when the underlying intensity patterns of the scene include ambiguities or repetitive patterns. An advantage of this method is that we can eliminate false marches and increase precision without anv search or sequentialfiltering.

This paper first defines a stereo algorithm based on the SSSD-in-inverse-distance and presents a marhemarical analysis to show how the algorithm can remove ambiguity and increase precision. Then, a few experimental results with real stereo images are presented to demonstrate the effectiveness of the algorithm

1 Introduction

Stereo is a useful technique for obtaining 3-D information from 2-D images in computer vision. In stereo matching. we measure the disparity d. which is the difference bctwccn the corresponding points of left and right imagcs. The disparity d is related to the distance z by

$$d = BF\frac{1}{z}. \qquad (1)$$

where B and F are baseline and focal length, respectively.

'This research was supported by the Defense Advanced Research Projects Agency (DOD) and monitored hy the Avionics Laboratory, Air Force Wright Aeronautical Laboratories. Aeronautical Systems Division (AFSC), Wright-PattersonAFB, Ohio 45433-6543 under Contract F33615-87-C-1499. ARPA Order No. 4976. Amendment 20. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied. of DARPA or the U.S. government.

'Current address: Information Systems Research Center, Canon Inc.890-12 Kashimada. Saiwai-ku. Kawasaki. 211. Japan.

This equation indicates that for the same distance the disparity is proportional to the baseline. or that the baseline length B acts as a magnification factor in measuring d in order to obtain z. That is. the estimated distance is more precise if we set the two cameras farther apart from each other. which means a longer baseline. A longer baseline, however, poses its own problem. Because a longer disparity range must be starched. matching is more difficult and thus there is a greater possibility of a false match. So there is a trade-off between precision and accuracy (correctness) in matching.

One of the most common methods to deal with the problem is a coarse-to-fine control strategy [1] - [5]. Matching is done at a low resolution to reduce false matches and then the result is used to limit the search range of marching at a higher resolution, where more precise disparity measurements are calculated. Using a coarse resolution. however. does not always remove false matches. This is especially true when there is inherent ambiguity in matching, such as a repeated pattern over a large pan of the scene (eg., a scene of a picket fence). Another approach to remove false matches and to increase precision is to use multiple images. especially a sequence of densely sampled images along a camera path [6] - [9]. A shon baseline between a pair of consecutive images makes the matching or tracking of features easy, while the structure imposed by the camera motion allows integration of the possibly noisy individual measurements into a precise estimate. The integration has been performed either by exploiting constraints on the EPI [6][7] or by a sequential Kalman filtering technique [8][9].

The stereo matching method presented in this paper belongs to the second approach use of multiple images with different baselines obtained by a lateral displacement of a camera. The matching technique. however. is based on the idea that global mismatches can be reduced by adding the sum of squared-diffcrcncc (SSD) values from multiple stereo pairs. That is. the SSD values are computed first for each pair of stereo images. We represent the SSD values with respect to the inverse distance $1/z$ (rather than the disparity d, as is usually done). The resulting SSD functions from all stereo pairs are added together to produce the sum of SSDs. which we call SSSD-in-inverse-distance. We show that the SSSD-in-inverse-distance function exhibits a unique and clear minimum at the correct matching position even when the underlying intensity patterns of the scene include ambiguities or repetitive patterns.

There have been stereo techniques that use multiple image pairs taken by cameras which are arranged along a line [10][11][12], in the form of a triangle [13][14][15] (called trinocular stereo), or in the other formation [16]. However, all of these techniques. except [10] and [16], decide candidate points for correspondence in each image pair and then search for the correct combinations of correspondences among them using the geometrical consistencies that they must satisfy. Since the intermediate decisions on correspondences are inherently noisy. ambiguous and multiple. finding the correct combinations requires sophisticated consistency checks and search or filtering. In contrast. our method does hot make my

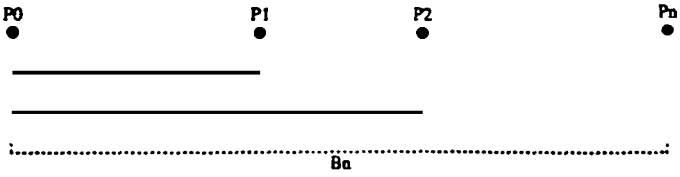


Figure 1: Camera positions for stereo

decisions about the correspondences in each stereo image pair. instead, it simply accumulates the measures of matching (SSDs) from all the stereo pairs into a single evaluation function. ie.. SSSD-in-inverse-distance, and then obtains one corresponding point from it In other words, our method integrates *evidence* for a final decision. rather than filtering intermediate *decisions*. In this sense, Tsai [16] employed strategy very similar to ours: he used multiple images to sharpen the peaks of his overall similarity measures. which he called JMM and WVM. However. the relationship between the improvement of the similarity measures and the camera baseline arrangement was not analyzed, nor was the method tested with real imagery. In this paper we show both mathematical analysis and experimental results with real indoor and outdoor images. which demonstrate how the SSSD-in-inversedistance function based on multiple image pairs from different baselines can greatly reduce false matches, while improving precision.

In the next section we present the method mathematically and show how ambiguity can be removed and prccision increased by the method. Section 3 provides a few experimental results with real stereo images to demonstrate the effectiveness of the algorithm. Section 4 presents conclusions.

2 Mathematical Analysis

The essence of stereo matching is. given a point in one image. to find in another image the corresponding point. such that the two points are the projections of the Same physical point in space. This task usually requires some criterion to measure similarity between images. The sum of squared differences (SSD) of the intensity values (or values of preprocessed images. such as bandpass filtered images) over a window is the simplest and most effective criterion. In this section, we define the sum of SSD with respect to the inverse distance (SSSD-in-inversedistance) for multiple-baseline stereo. and mathematically show its advantage in removing ambiguity and increasing precision. For this analysis, we use 1-D stereo intensity signals, but the extension to two dimensional images is straightforward.

2.1 SSD Function

Suppose that we have cameras at positions $P_0, P_1, \ldots, P_n$ along a line with their optical axes perpendicular to the line and a resulting set of stereo pairs with baselines $B_1, B_2, \ldots, B_n$ as shown in figure 1. Let $f_0(x)$ and $f_i(x)$ be the image pair at the camera positions P_0 and P_i, respectively. Imagine a scene point Z whose distance is z. Its disparity $d_{r(i)}$ for the image pair taken from P_0 and P_i is

$$d_{r(i)} = \frac{B_i F}{z}. \tag{2}$$

We model the image intensity functions $f_0(x)$ and $f_i(x)$ near the matching positions for Z as

$$\begin{aligned} f_0(x) &= f(x) + n_0(x) \\ f_i(x) &= f(x - d_{r(i)}) + n_i(x), \end{aligned} \tag{3}$$

assuming constant distance near Z and independent Gaussian white noise such that

$$n_0(x), n_i(x) \sim N(0, \sigma_n^2). \tag{4}$$

The SSD value $e_{d(i)}$ over a window W at a pixel position x of image $f_0(x)$ for the candidate disparity $d_{(i)}$ is defined as

$$e_{d(i)}(x, d_{(i)}) \equiv \sum_{j \in W} (f_0(x + j) - f_i(x + d_{(i)} + j))^2, \tag{5}$$

where the $\sum_{j \in W}$ means summation over the window. The $d_{(i)}$ that gives a minimum of $e_{d(i)}(x, d_{(i)})$ is determined as the estimate of the disparity at x. Since the SSD measurement $e_{d(i)}(x, d_{(i)})$ is a random variable, we will compute its expected value in order to analyze its behavior.

$$\begin{aligned} & E[e_{d(i)}(x, d_{(i)})] \\ &= E\left[\sum_{j \in W} (f(x + j) - f(x + d_{(i)} - d_{r(i)} + j) \right. \\ &\quad \left. + n_0(x + j) - n_i(x + d_{(i)} + j))^2\right] \\ &= \sum_{j \in W} (f(x + j) - f(x + d_{(i)} - d_{r(i)} + j))^2 + 2N_w\sigma_n^2. \end{aligned} \tag{6}$$

where N_w is the number of the points within the window. For the rest of the paper. $E\{\}$ denotes the expected value of a random variable. In deriving the above equation. we have assumed that $d_{r(i)}$ is constant over the window. Equation (6) says that naturally the SSD function $e_{d(i)}(x, d_{(i)})$ is expected to take a minimum when $d_{(i)} = d_{r(i)}$, i.e., at the right disparity.

Let us examine how the SSD function $e_{d(i)}(x, d_{(i)})$ behaves when there is ambiguity in the underlying intensity function. Suppose that the intensity signal $f(x)$ has the same pattern around pixel positions x and $x + a$.

$$f(x + j) = f(x + a + j), \qquad j \in W \tag{7}$$

where a $\neq 0$ is a constant Then. from equation (6)

$$E[e_{d(i)}(x, d_{r(i)})] = E[e_{d(i)}(x, d_{r(i)} + a)] = 2N_w\sigma_n^2. \tag{8}$$

This means that ambiguity is expected in matching in terms of positions of minimum SSD values. Moreover. the false match at $d_{r(i)} + a$ appears in exactly the Same way for all i; it is separated from the correct match by a for all the stereo pairs. Using multiple baselines does not help to disambiguate.

2.2 SSD with respect to Inverse Distance

Now, let us introduce the *inverse distance* ζ such that

$$\zeta = \frac{1}{z}. \tag{9}$$

>From equation and (2),

$$d_{r(i)} = B_i F \zeta_r \tag{10}$$

$$d_{(i)} = B_i F \zeta, \tag{11}$$

where ζ_r and ζ are the real and the candidate inverse distance. respectively. Substituting equation (11) into (5), we hive the SSD with respect to the inverse distance,

$$e_{\zeta(i)}(x, \zeta) \equiv \sum_{j \in W} (f_0(x + j) - f_i(x + B_i F \zeta + j))^2, \tag{12}$$

at position x for a candidate inverse distance ζ. Its expected value is

$$E[e_{\zeta(i)}(x,\zeta)] = \sum_{j\in W}(f(x+j)-f(x+B_iF(\zeta-\zeta_r)+j))^2+2N_w\sigma_n^2. \quad (13)$$

Finally, we define a new evaluation function $e_{\zeta(12\cdots n)}(x,\zeta)$, the sum of SSD functions with respect to the inverse distance (SSSD-in-inverse-distance) for multiple stereo pairs. It is obtained by adding the SSD functions $e_{\zeta(i)}(x,\zeta)$ for individual stereo pairs:

$$e_{\zeta(12\cdots n)}(x,\zeta) = \sum_{i=1}^{n} e_{\zeta(i)}(x,\zeta). \quad (14)$$

Its expected value is

$$\begin{aligned} E[e_{\zeta(12\cdots n)}(x,\zeta)] &= \sum_{i=1}^{n} E[e_{\zeta(i)}(x,\zeta)] \\ &= \sum_{i=1}^{n}\sum_{j\in W}(f(x+j)-f(x+B_iF(\zeta-\zeta_r)+j))^2 \\ &\quad +2nN_w\sigma_n^2. \end{aligned} \quad (15)$$

In the next three subsections. we will analyze the characteristics of these evaluation functions to see how ambiguity is removed and precision is improved.

2.3 Elimination of Ambiguity (1)

As before, suppose the underlying intensity pattern $f(x)$ has the same pattern around x and $x+a$ (equation (7)). Then. according to equation (13). we have

$$E[e_{\zeta(i)}(x,\zeta_r)] = E[e_{\zeta(i)}(x,\zeta_r+\tfrac{a}{B_iF})] = 2N_w\sigma_n^2. \quad (16)$$

We still have an ambiguity; a minimum is expected at a false inverse distance $\zeta_f = \zeta_r + \frac{a}{B_iF}$. However. an important point to be observed here is that this minimum for the false inverse distance ζ_f changes its position as the baseline B_i changes, while the minimum for the correct inverse distance ζ_r does not. This is the property that the new evaluation function. the SSSD-in-inversedistance (14). exploits to eliminate the ambiguity. For example, suppose we use two baselines B_1 and B_2 ($B_1 \neq B_2$). >From equation (15)

$$\begin{aligned} &E[e_{\zeta(12)}(x,\zeta)] \\ &= \sum_{j\in W}(f(x+j)-f(x+B_1F(\zeta-\zeta_r)+j))^2 \\ &+ \sum_{j\in W}(f(x+j)-f(x+B_2F(\zeta-\zeta_r)+j))^2 \\ &+ 4N_w\sigma_n^2. \end{aligned} \quad (17)$$

We can prove that

$$E[e_{\zeta(12)}(x,\zeta)] > 4N_w\sigma_n^2 = E[e_{\zeta(12)}(x,\zeta_r)] \quad \text{for } \zeta \neq \zeta_r. \quad (18)$$

(refer to appendix A) In words. $e_{\zeta(12)}(x,\zeta)$ is expected to have the smallest value at the correct ζ_r. That is. the ambiguity is likely to be eliminated by use of the new evaluation function with two different baselines.

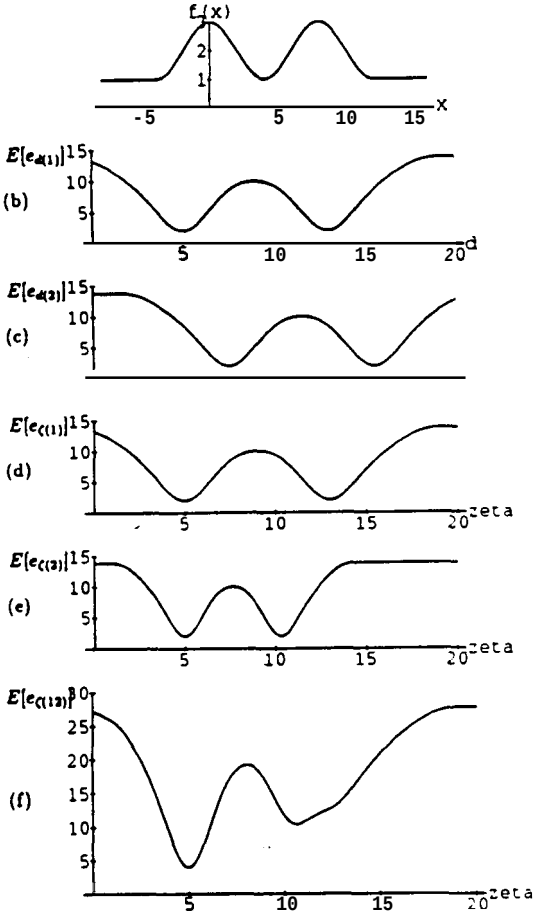


Figure 2: Expected values of evaluation functions: (a) Underlying function; (b) $E[e_{d(1)}]$; (c) $E[e_{d(2)}]$; (d) $E[e_{\zeta(1)}]$; (e) $E[e_{\zeta(2)}]$; (f) $E[e_{\zeta(12)}]$

We can illustrate this using synthesized data. Suppose the point whose distance we want to determine is at $x = 0$ and the underlying function $f(x)$ is given by

$$f(x) = \begin{cases} \cos(\frac{\pi}{4}x)+2 & \text{if } -4 < x < 12 \\ 1 & \text{if } x \leq -4 \text{ or } 12 \leq x. \end{cases} \quad (19)$$

Figure 2 (a) shows a plot of $f(x)$. Assuming that $d_{r(1)} = 5$. $\sigma_n^2 = 0.2$, and the window size is 5, the expected values of the SSD function $e_{d(1)}(x, d_{(1)})$ are as shown in figure 2 (b). We see that there is an ambiguity: the minima occur at the correct match $d_{(1)} = 5$ and at the false match $d_{(1)} = 13$. Which match will be selected will depend on the noise, search range. and search strategy. Now suppose we have a longer baseline B_2 such that $\frac{B_2}{B_1} = 1.5$. >From equations (6) and (10), we obtain $E[e_{d(2)}]$ as shown in figure 2 (c). Again we encounter an ambiguity, and the separation of the two minima is the same.

Now let us evaluate the SSD values with respect to the inverse distance ζ rather than the disparity d by using equations (12) through (15). The expected values of the SSD measurements $E[e_{\zeta(1)}]$ and $E[e_{\zeta(2)}]$ with baselines B_1 and B_2 are shown in figures 2 (d) and (e). respectively (the plot is normalized such that $B_1F = 1$). Note that the minima at the correct inverse distance ($\zeta = 5$) does not move. while the minima for the false

(a) (b)

Figure 3: "Town" data set: (a) Image0; (b) Image9

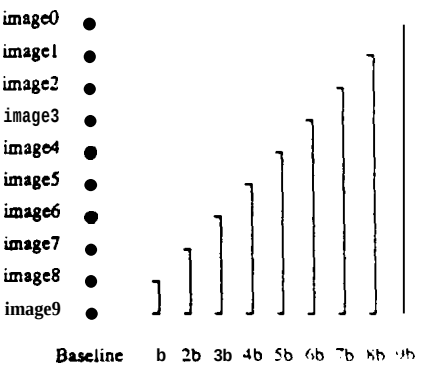


Figure 4: "Town" data set image sequence

match changes its position as the baseline changes. When the two functions are added to produce the SSSD-in-inverse-distance, its expected values $E[e_{\zeta(12)}]$ are shown in figure 2 (f). We can see that the ambiguity has been reduced because the SSSD-in-inverse-distance has a smaller value at the correct match position than at the false match.

24 Elimination of Ambiguity (2)

An extreme case of ambiguity occurs when the underlying function $f(x)$ is a periodic function. like a scene of a picket fence. We can show that this ambiguity can also be eliminated.

Let $f(x)$ be a periodic function with period T. Then. each $e_{\zeta(i)}(x,\zeta)$ is expected to be a periodic function of ζ with the period $\frac{T}{B_i F}$. This means that there will be multiple minima of $e_{\zeta(i)}(x,\zeta)$ (i.e., ambiguity in matching) at intervals of $\frac{T}{B_i F}$ in ζ. When we use two baselines and add their SSD values. the resulting $e_{\zeta(12)}(x,\zeta)$ will be still a periodic function of ζ, hut its period T_{12} is increased to

$$T_{12} = LCM\left(\frac{T}{B_1 F}, \frac{T}{B_2 F}\right) \qquad (20)$$

where $LCM()$ denotes Least Common Multiple. That is. the period of the expected value of the new evaluation function can be made longer than that of the individual stcrco pairs. Furthermore. it can be controlled by choosing the basclines B_1 and B_2 appropriately so that the expected value of the evaluation function has only one minimum within the search range. This means that using multiple-baseline stereo pairs simultaneously can eliminate ambiguity, although each individual baseline stcrco may suffer from ambiguity.

We illustrate this by using real stereo imagcs. Figurc 3(a) shows an image of a sample scene. A i the top of the scene thcre is a grid board whose intensity function is nearly periodic. Wc took ten images of this scene by shifting the camera vcnically as in figure 4. The actual distance between consecutive camera positions is 0.05 inches. Let this distance be b. Figure 3 shows the first and the last images of the sequence. We selected a point x within the repetitive grid board area in image9. The SSD values $e_{\zeta(1)}(x,\zeta)$ over 5-by-5-pixel windows are plotted for various baseline sterco pairs in figure 5. The horizontal axis of all the plots is the inverse distance. normalized such that $8bF = 1$. Figure 5 illustrates the trade-off between precision and ambiguity in terms of baselines. That is. fora shorter baseline. there are fewer minima (i.e. less ambiguity). but the SSD curve is Ratter (i.e. less precise localization). On ihc other hand. for a longer baseline. there are more minima (i.e. mort ambiguity), but the curve near the minimum is sharper; that is. the estimated distance is more precise if we can find the correct one.

Now. Ictus take two stereo image pairs: one with $B = 5b$ and the other with $B = 8b$. In figure 6. the dashed curve and the dotted curve show the SSD for $B = 5b$ and $B = 8b$, respectively. Let us suppose the search range goes from 0 to 20 in the horizontal axis, which in this case corresponds to 12 io ∞ inches in distance. Though the SSD values take a minimum at the correct answer near $\zeta = 5$, there are also other minima for both cases. The solid curve shows the evaluation function for the multiple-haw lint stereo, which is the sum of the dashed curve and the dotted curve. The solid curve shows only one clear minimum; that is. the ambiguity is resolved.

So far, we have considered using only two stcrco pairs. We can easily extend the idea to multiple-baseline stereo which uses more than two stereo pairs. Corresponding to equation (20), the period of $E[e_{\zeta(12\cdots n)}(x,\zeta)]$ becomes

$$T_{12,\ldots,n} = LCM\left(\frac{T}{B_1 F}, \frac{T}{B_2 F}, \ldots, \frac{T}{B_n F}\right) \qquad (21)$$

where $B_1, B_2, \ldots, B_n$ are baselines for each stereo pair.

We will demonstrate how the ambiguity can be further reduced by increasing the number of stereo pairs. >From the data of figure 4. we first choose image1 and image9 as a long baseline stereo pair. ie. (1) $B = 8b$. Then, we increase the number of sterco pairs by dividing the baseline between image1 and image9. i.e. (2) $B = 4b$ and $8b$. (3) $B = 2b$, $4b$, $6b$ and $8b$. (4) $B = b$, $2b$, $3b$, $4b$, $5b$, $6b$, $7b$ and $8b$. Figure 7 demonstrates that the SSSDs-in-inverse-distance shows the minimum at the correct position more clearly as more stereo pairs are used.

25 Precision

We have shown that ambiguities can be resolved by using the SSSD-in-inverse-distance computed from multiple baseline stereo pairs. The technique also increases precision in estimating the true inverse distance. We can show this by analyzing the statistical characteristics of the evaluation functions near the correct match.

>From equations (3), (10).and (12), we have

$$e_{\zeta(1)}(x,\zeta) = \sum_{j\in W}(f(x+j) - f(x+B_1 F(\zeta-\zeta_r)+j) + n_0(x+j) - n_1(x+B_1 F\zeta+j))^2 \qquad (22)$$

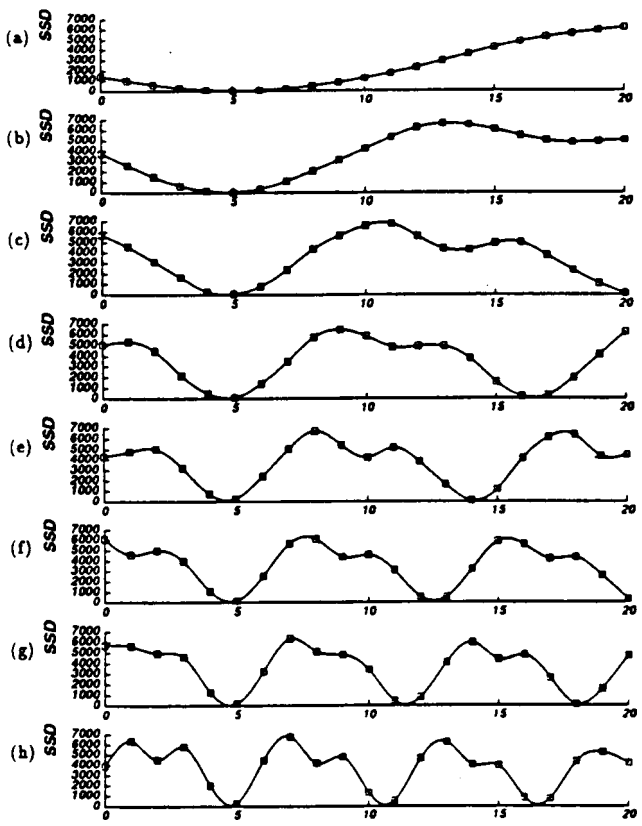


Figure 5: SSD values vs. inverse depth: (a) B = b; (b) B = $2b$; (c) B = 3b; (d) B = $4b$; (e) B = $5b$; (f) B = 66; (g) B = 7b; (h) B = $8b$. The horizontal axis is normalized such that $8bF = 1$.

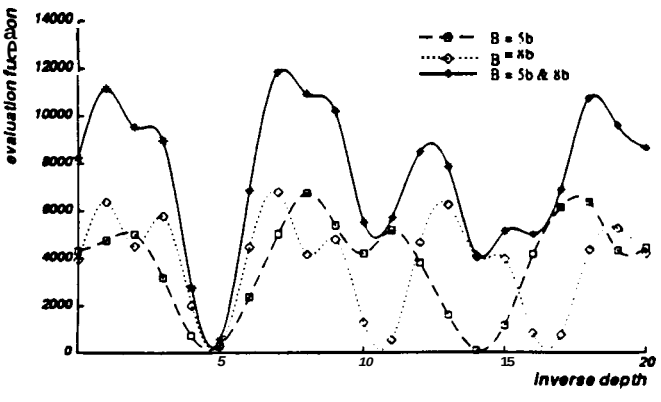


Figure 6: Combining two stereo pairs with different baselines

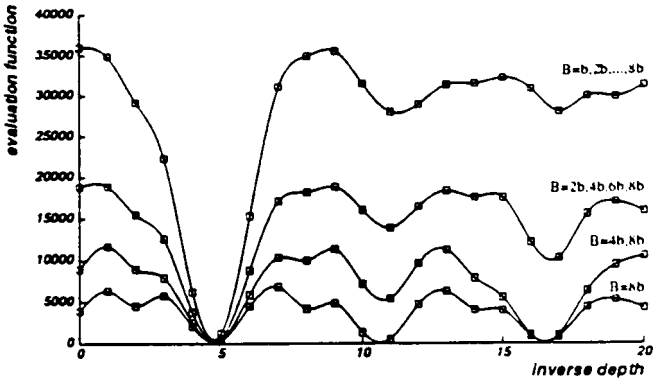


Figure 7: Combining multiple baseline stereo pairs

By taking the Taylor expansion about $\zeta = \zeta_r$ up to the linear terms, we obtain

$$f(x+B_iF(\zeta-\zeta_r)+j) \approx f(x+j)+B_iF(\zeta-\zeta_r)f'(x+j). \quad (23)$$

Substituting this into equation (22), we can approximate $e_{\zeta(i)}(x,\zeta)$ near ζ_r by a quadratic form of ζ:

$$\begin{aligned} & e_{\zeta(i)}(x,\zeta) \\ \approx & \sum_{j\in W}(-B_iF(\zeta-\zeta_r)f'(x+j) \\ & +n_0(x+j)-n_i(x+B_iF\zeta+j))^2 \\ = & B_i^2F^2a(x)(\zeta-\zeta_r)^2+2B_iFb_i(x)(\zeta-\zeta_r)+c_i(x). \end{aligned} \quad (24)$$

where

$$a(x) = \sum_{JEW}(f'(x+j))^2 \quad (25)$$

$$b_i(x) = \sum_{JEW} f'(x+j)(n_i(x+B_iF\zeta+j)-n_0(x+j)) \quad (26)$$

$$c_i(x) = \sum_{j\in W}(n_i(x+B_iF\zeta+j)-n_0(x+j))^2. \quad (27)$$

The estimated inverse distance $\zeta_{r(i)}$ is the value ζ that makes equation (24) minimum:

$$\zeta_{r(i)} = \zeta_r - \frac{b_i(x)}{B_iFa(x)}. \quad (28)$$

Since $E[b_i(x)] = 0$, the expected value of the estimate $\zeta_{r(i)}$ is the correct value ζ_r, but it varies due to the noise. The variance of this estimate is:

$$\begin{aligned} Var(\zeta_{r(i)}) = & \frac{Var(b_i(x))}{B_i^2F^2(a(x))^2} \\ = & \frac{2\sigma_n^2}{B_i^2F^2a(x)}. \end{aligned} \quad (29)$$

Basically, this equation states that for the same mount of image noise σ_n^2, the variance is smaller (the estimate is more precise) as the baseline B_i is longer, or as the variation of intensity signal, $a(x)$, is larger.

We can follow the Same analysis for $e_{\zeta(12\cdots n)}(x,\zeta)$ of (14), the new evaluation function with multiple baselines. Near ζ_r, it is

$$e_{\zeta(12\cdots n)}(x,\zeta) \approx \left(\sum_{i=1}^{n} B_i^2\right) F^2 a(x)(\zeta-\zeta_r)^2 + 2F\left(\sum_{i=1}^{n} B_i b_i(x)\right)(\zeta-\zeta_r) + \sum_{i=1}^{n} c_i(x). \quad (30)$$

The variance of the estimated inverse distance $\xi_{r(12\cdots n)}$ that minimizes this function is

$$Var(\xi_{r(12\cdots n)}) = \frac{2\sigma_n^2}{\left(\sum_{i=1}^{n} B_i^2\right) F^2 a(x)}. \quad (31)$$

>From equations (29) and (31), we see that

$$\frac{1}{Var(\xi_{r(12\cdots n)})} = \sum_{i=1}^{n} \frac{1}{Var(\xi_{r(i)})}. \quad (32)$$

The inverse of the variance represents the precision of the estimate. Therefore, equation (32) means that by using the SSSD-in-invcrsc-distance with multiple baseline stereo pairs. the estimate becomes more precise. We can confirm this characteristic in figures 6 and 7 by observing that the curve around the corrcct inverse distance becomes sharper as more baselines are used.

3 Experimental Results

This section presents experimental results of the multiple-baseline stereo based on SSSD-in-invcrsc-distance with real 2D images. A complete description of the algorithm is included in Appendix B.

The first result is for the "Town" data set that we showcd in figure 3. Figures 8 (a) and (b) are the distance map and its isometric plot with a short baseline. $B = 3b$. The result with a single long baseline, $B = 9b$. is shown in figure 9. Comparing these two results. we observe that the distance map computed by using the long baseline is smoother on flat surfaces, i.e., more precise, but has gross errors in matching at the top of the scene because of the repeated pattern. These results illustrate the trade-off between ambiguity and precision. Figure 10, on the other hand. shows the distance map and its isometric plot obtained by the new algorithm using three different basclines. 3b, 6b. and 9b. For comparison. the corresponding oblique view of the scene is shown in figure 11. We can note that the computed distance map is less ambiguous and more precise than those of the single-baseline stereo.

Figure 12 shows another data set used for our experiment. Figures 13 and 14 compare the distance maps computed from the shon baseline stereo and the long baseline stereo: the longer baseline is five times longer than the short one. For comparison. the actual oblique view roughly corresponding to the isometric plot is shown in figure 15. Though no repetitive patterns are apparent in the images, we can still observe gross errors in the distance map obtained with the long baseline due to false matching. In contrast, the rcsult from the multiple-baseline stereo shown in figure 16 dcmonstratcs both the advantage of unambiguous matching with a short basclinc and that of precise matching with a long baseline.

4 Conclusions

In this paper. we have presented a new stereo matching method which uses multiple baseline stcrco pairs. This method can overcome the trade-off between precision and accuracy (avoidance of false matches) in stereo. The method is rather straightforward: we represent the SSD values for individual stereo pairs as a function of the inverse distance. and add those functions. The resulting function, the SSSD-in-inverse-distance, exhibits an unambiguous and sharper minimum at the correct matching position. As a rcsult there is no need for search or sequential estimation proccdurcs.

The key idea of the method is to relate SSD values to the invcrsc distance rather than the disparity. As an afterthought. this idea is natural. Whereas disparity is a function of the bascline. thcrr is only one true (inverse) distance for each pixel position for all of the stereo pairs. Therefore there must be a single minimum ior the SSD values when they are summed and plotted with respect to the inverse distance. We have shown the advantage of the proposed method in removing ambiguity and improving prccision by analytical and experimental results.

Acknowledgment

The authors would like to thank John Krumm for his useful comments on this paper. Keith Gremban. Jim Rehg and Carol Novak have read the manuscript and improved its readability substantially.

A SSSD-in-inverse-distance for Ambiguous Pattern

Proposition: Suppose that there are two and only two repetitions of the same pattern around positions x and $x + a$ whcrc $a \neq 0$ is a constant. That is. for $j \in W$

$$f(x+j) = f(\xi+j), \quad \text{if and only if } \xi = x \text{ or } \xi = x+a. \quad (33)$$

Then. if $B_1 \neq B_2$, for $\forall \zeta, \zeta \neq \zeta_r$,

$$\begin{aligned} & E[e_{\zeta(12)}(x,\zeta)] \\ = & \sum_{j\in W} (f(x+j) - f(x+B_1F(\zeta-\zeta_r)+j))^2 \\ + & \sum_{j\in W} (f(x+j) - f(x+B_2F(\zeta-\zeta_r)+j))^2 + 4N_w\sigma_n^2 \\ > & 4N_w\sigma_n^2 = E[e_{\zeta(12)}(x,\zeta_r)]. \end{aligned} \quad (34)$$

Proof: Tentatively suppose that for $\exists \zeta_f, \zeta_f \neq \zeta_r$,

$$\begin{aligned} & \sum_{j\in W} (f(x+j) - f(x+B_1F(\zeta_f-\zeta_r)+j))^2 \\ + & \sum_{j\in W} (f(x+j) - f(x+B_2F(\zeta_f-\zeta_r)+j))^2 \\ = & 0. \end{aligned} \quad (35)$$

Then. it must be the case that

$$\begin{aligned} f(x+j) &= f(x+a_1+j) \\ \text{and } f(x+j) &= f(x+a_2+j), \end{aligned} \quad (36)$$

for $j \in W$. where

$$\begin{aligned} a_1 &= B_1F(\zeta_f-\zeta_r) \\ a_2 &= B_2F(\zeta_f-\zeta_r). \end{aligned}$$

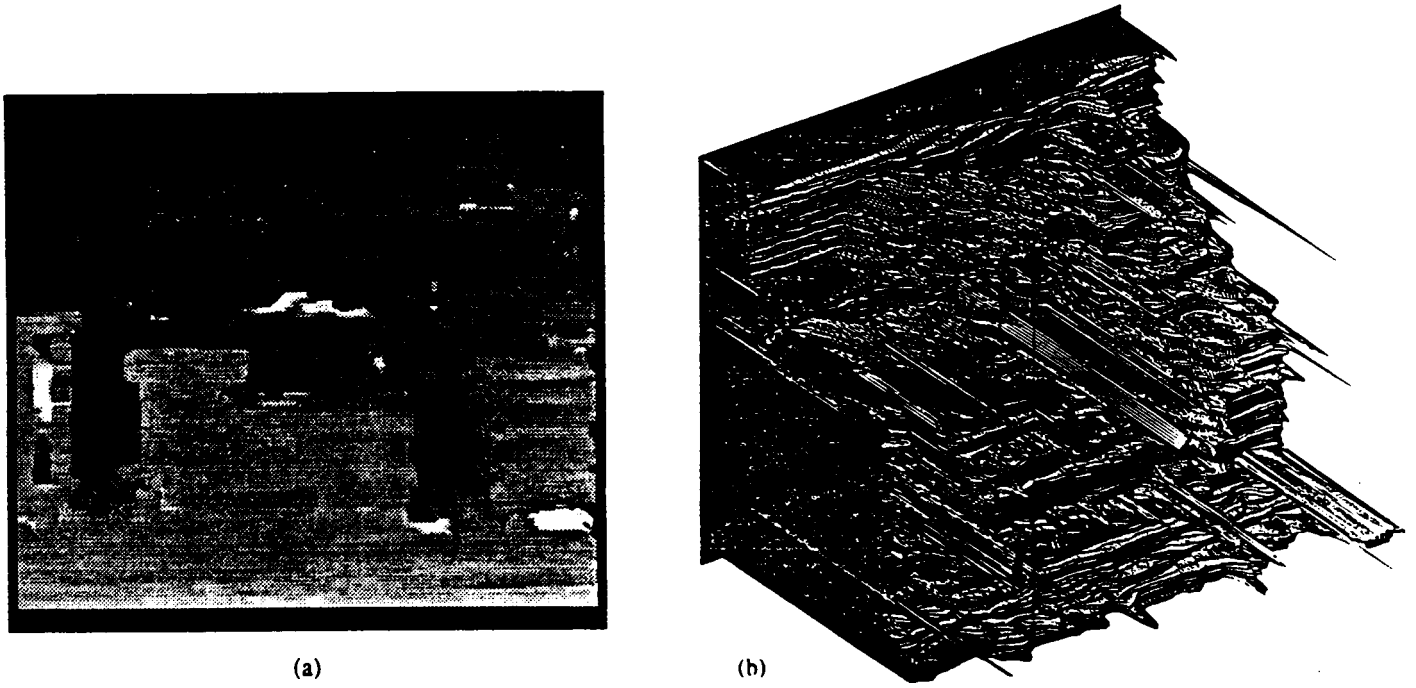
Since $B_1 \neq B_2$ and $\zeta_r \neq \zeta_f$,

$$a_1 \neq a_2. \quad (37)$$

So. we have

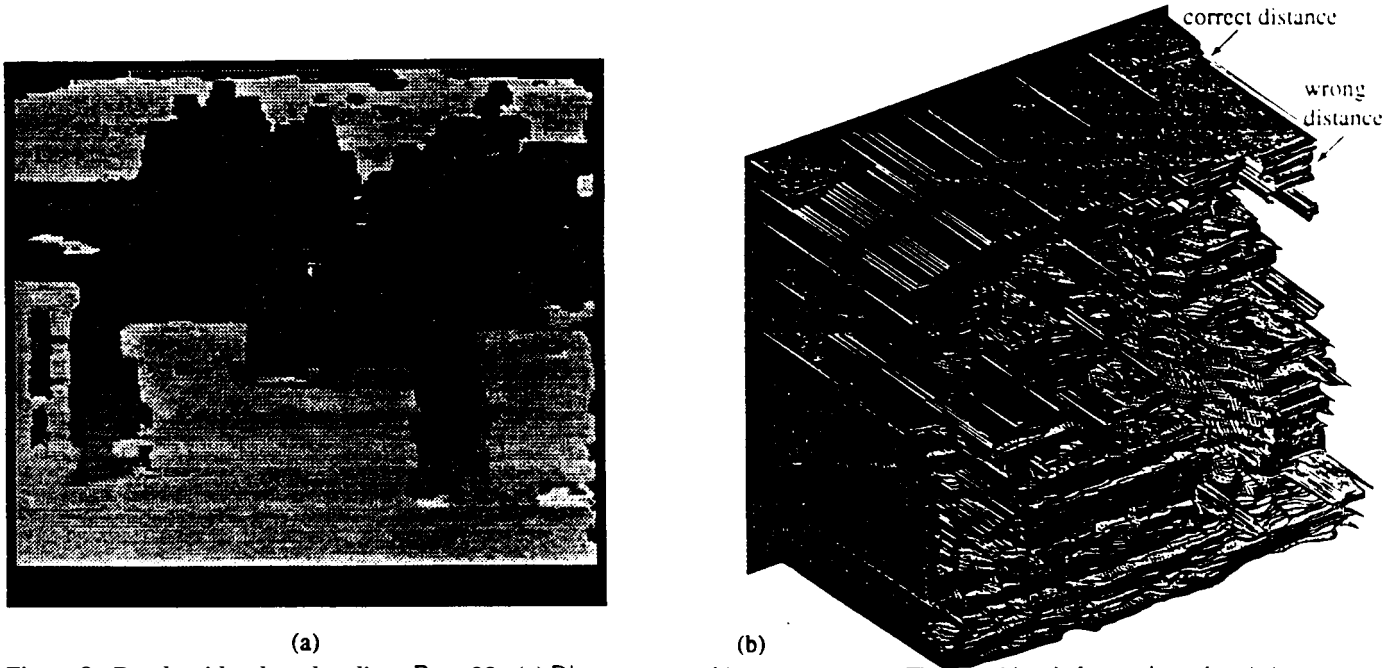
$$f(x+j) = f(\xi+j), \quad \text{for } \xi = x,\ x+a_1, \text{ or } x+a_2. \quad (38)$$

Since this contradicts assumption (33), equation (35) does not hold. Its lcft hand side must be positive. Hence (34) holds.



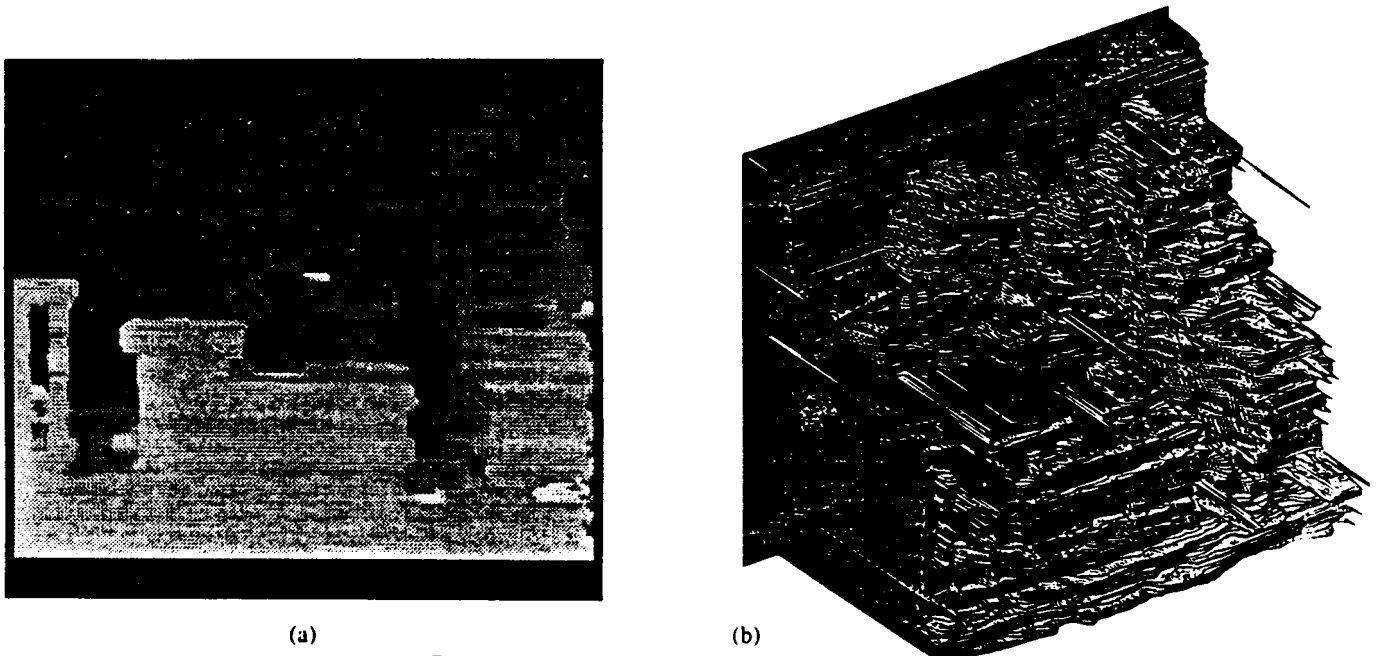
(a) (b)

Figure 8: Result with a short baseline, $E = 36$: (a) Distance map; (b) Isometric plot of the distance map from the upper left corner. The matching is mostly correct. but very noisy.



(a) (b)

Figure 9: Result with a long baseline, $B = 96$: (a) Distance map; (b) Isometric plot. The matching is less noisy when it is correct. However, there are many gross mistakes. especially in the top of the image where, due to a repetitive pattern. the matching is completely wrong.



(a) (b)

Figure 10: Result with multiple baselines, $B = 3b, 6b$, and $9b$: (a) Distance map; (b) Isometric plot. Compared with figures 8(b) and 9(b), we see that the distance map is less noisy and that gross errors have been removed.

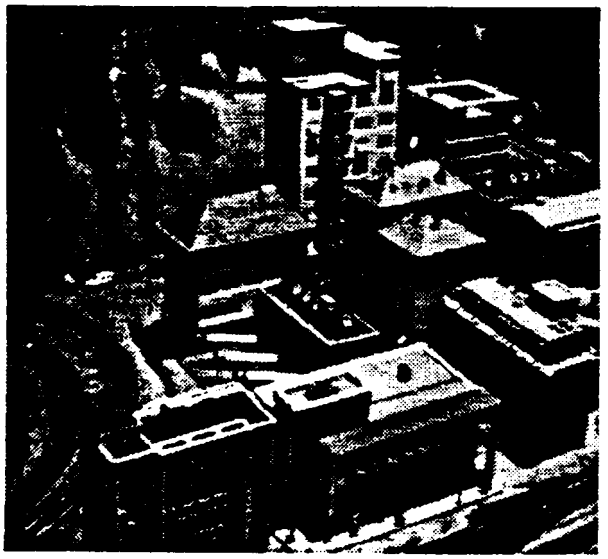
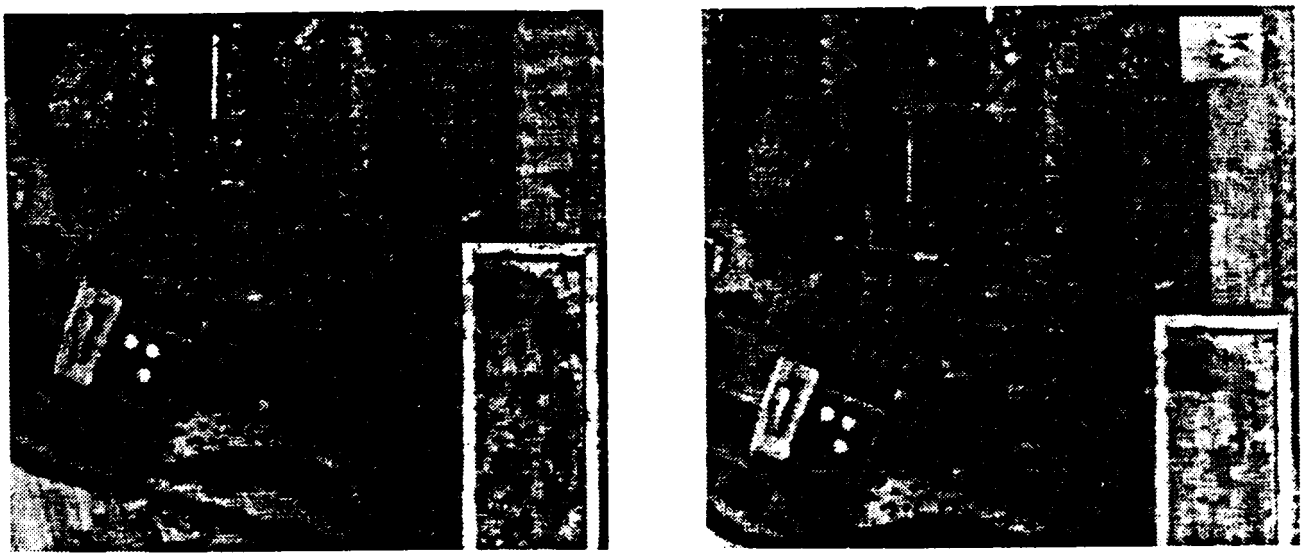
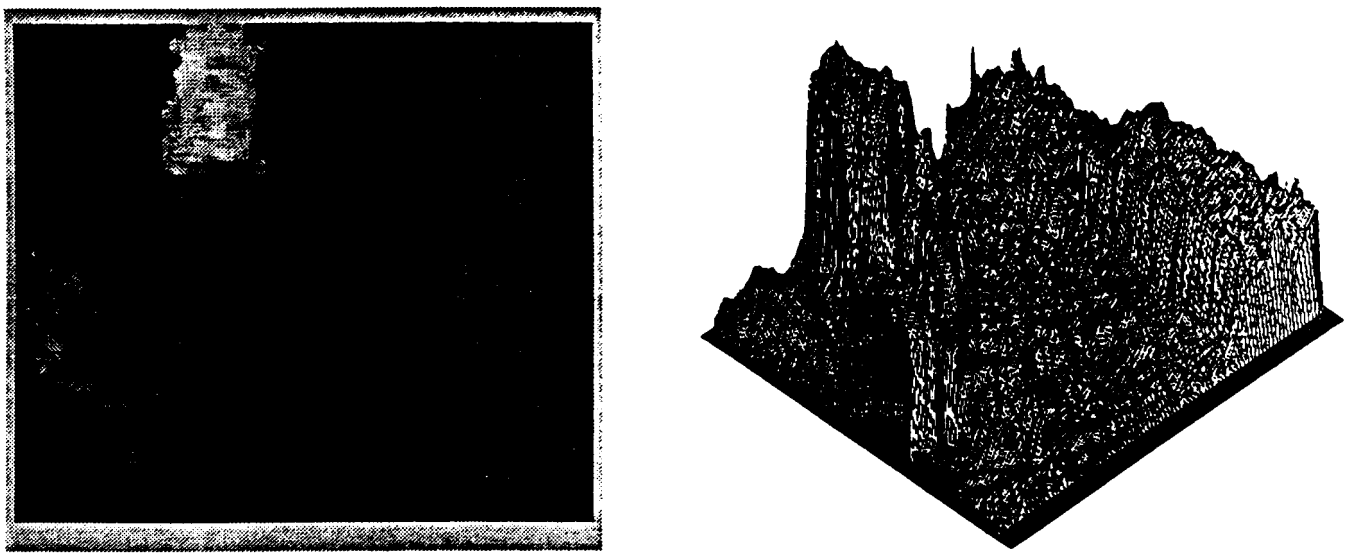


Figure 11: Oblique view



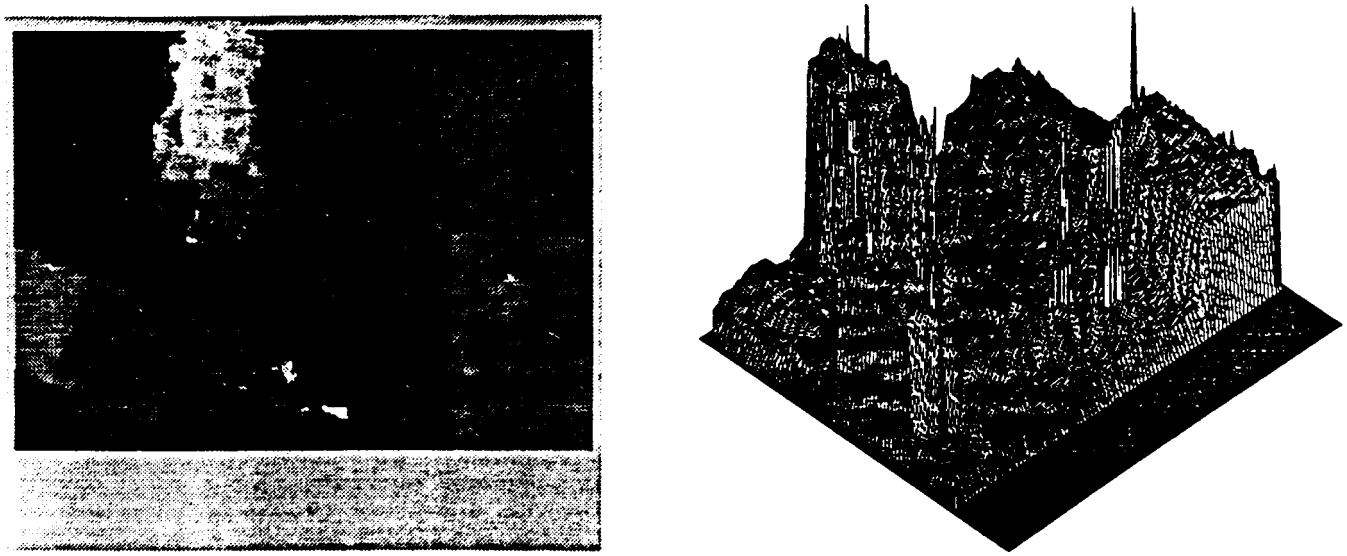
(a) (b)

Figure 12: "Coal mine' data set. long-baseline pair



(a) (b)

Figure 13: Result with a short baseline: (a) Distance map: (b) Isometric plot of the distance map viewed from the lower left corner



(a) (b)

Figurc 14: Result **with a** long baseline: (a) Distance **map: (b) Isometric** plot

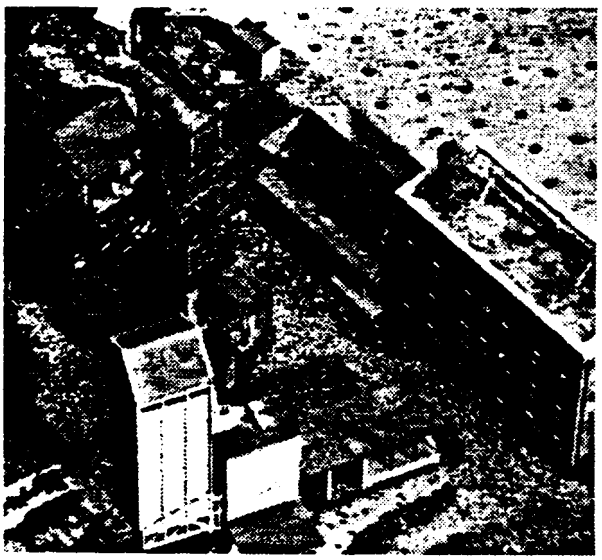
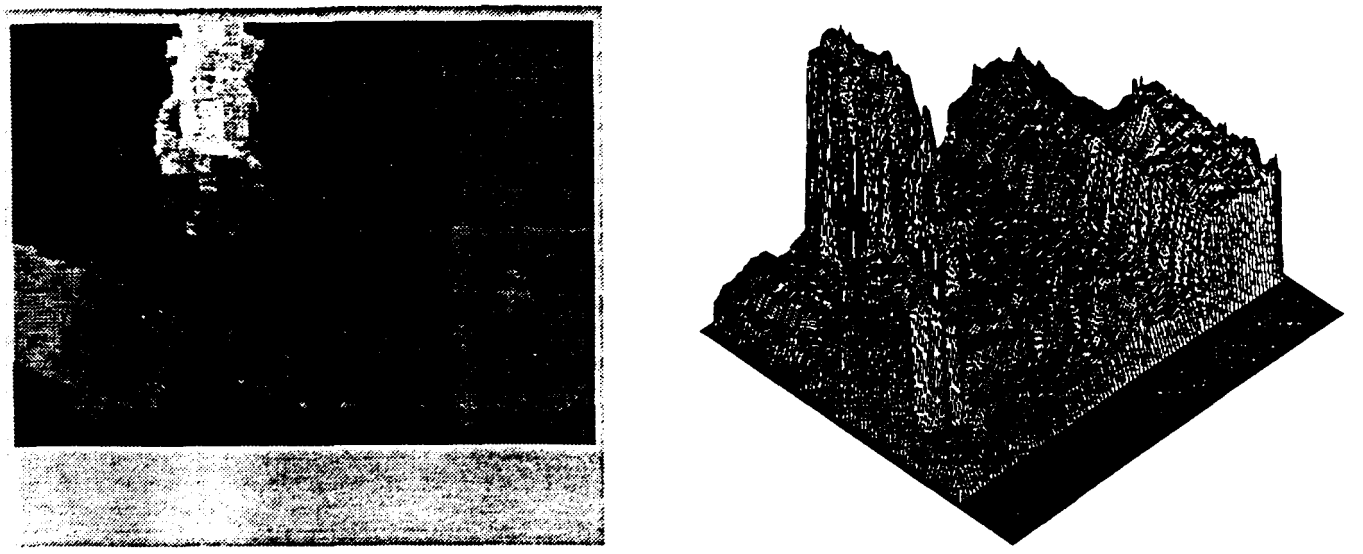


Figure 15: **Oblique view**



(a) (b)

Figure 16: Multiple baselines: (a) Distance map; (b) Isometric plot

B Multiple-Baseline Stereo Algorithm

We present a complete description of the stereo algorithm using multiple-hascline stereo pairs. The task is. given n stereo pairs. tind the ζ that minimizes the SSSD-in-inverse-distance function.

$$SSSD(x,\zeta) = \sum_{i=1}^{n} \sum_{j \in W} (f_0(x+j) - f_i(x + B_i F \zeta + j))^2. \quad (39)$$

We will perform this task in two steps: one at pixel resolution hy minimum detection and the other at sub-pixel resolution hy iterative estimation.

Minimum of SSSD at Pixel Resolution

For convenience. instead of using the inverse distance. we normalize the disparity values of individual stereo pairs with different baselines to the corresponding values for the largest baseline. Suppose $B_1 < B_2 < \ldots < B_n$. We define the baseline ratio R, such that

$$R_i = \frac{B_i}{B_n}. \quad (40)$$

Then.

$$B_i F \zeta = R_i B_n F \zeta = R_i d_{(n)}. \quad (41)$$

where $d_{(n)}$ is the disparity for the stereo pair with baseline B_n. Substituting this into equation (39),

$$SSSD(x, d_{(n)}) = \sum_{i=1}^{n} \sum_{j \in W} (f_0(x+j) - f_i(x + R_i d_{(n)} + j))^2. \quad (42)$$

We compute the SSSD function for a range of disparity values at the pixel resolution. and identify the disparity that gives the minimum. Note that pixel resolution for the image pair with the longest baseline (B_n) requires calculation of SSD values at sub-pixel resolution for other shorter hascline stereo pairs.

Iterative Estimation at Sub-pixel Resolution

Once we obtain disparity at pixel resolution for the longest baseline stereo. we improve the disparity estimate to sub-pixel resolution hy an iterative algorithm presented in [12][17]. For this iterative estimation. we use only the image pair $f_0(x)$ and $f_n(x)$ with the longest baseline. This is due to a few reasons. First. since the pixel-level estimate was obtained by using the SSSD-in-inverse-distance. the ambiguity has been eliminated and only improvement of precision is intended at this stage. Second. using only the longest-baseline image pair reduces the computational requirement for SSD calculation by a factor of n, and yet does not degrade precision too significantly.

In the experiments shown in section 3. we used the following algorithm for sub-pixel estimation: Let $d_{0(n)}$ be the initial disparity estimate obtained at pixel resolution. Then, a more precise estimate is computed by calculating the following two quantities:

$$\Delta d_{(n)} = \frac{\sum_{j \in W} (f_0(x+j) - f_n(x + d_{0(n)} + j)) f_n'(x + d_{0(n)} + j)}{\sum_{j \in W} (f_n'(x + d_{0(n)} + j))^2} \quad (43)$$

$$\sigma^2_{\Delta d_{(n)}} = \frac{2\sigma_n^2}{\sum_{j \in W} (f_n'(x + d_{0(n)} + j))^2}, \quad (44)$$

The value $\Delta d_{(n)}$ is the estimate of the correction of the disparity to further minimize the SSD. and $\sigma^2_{\Delta d_{(n)}}$ is its variance. We iterate this procedure by replacing $d_{0(n)}$ by

$$d_{0(n)} \leftarrow d_{0(n)} + \Delta d_{(n)} \quad (45)$$

until the estimate converges or up to a certain maximum number of iterations.

References

[1] D. Marr and T. Poggio. A theory of human stereo vision. In *Proc. Roy. Soc. London.* volume vol. B 204, pages 301–328. 1979.

[2] W. E. L.Grimson. Computational experiments with a feature based stereo algorithm. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 7(1): 17-34. Jan. 1985.

[3] Stephen T. Barnard. Stochastic stereo matching over scale. *International Journal of Compurer Vision*, pages 17-32, 1989.

[4] M. J. Hannah. A system for digital stereo image matching. *Photogrammetric Engineering and Remote Sensing*, 55(12): 1765-1770, Dec. 1989

[5] Jer-sen Chen and Gerard Medioni. Parallel multiscale stereo matching using adaptive smoothing. in *ECCV'90*, pages99-103, 1990.

[6] R. C. Bolles. H. H. Baker. and D. H. Marimont. Epipolar-plane image analysis: An Approach to determining structure from motion. *Iniernarional Journal of Compurer Vision*, 1(1), 1987.

[7] Masanobu Yamamoto. The image sequence analysis oi three-dimensional dynamic scenes. Technical Report 893. Electrotechnical Laboratory - Agency of Industrial Science and Technology, Tsukuba, Ibaraki, Japan, May 1988.

[8] Larry Matthies. Richard Szeliski, and Takeo Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3:209-236, 1989.

[9] Joachim Heel. Dynamic motion vision. In *Proceedings of rhe DARPA Image Undestanding Workshop*, pages 702-713. Palo Alto. Ca. May 23-26 1989.

[10] B. Wilcox. Telerobotics and Mars Rover research at JPL. Lecture at CMU. Oct. 1987.

[11] Hans P. Moravec. Visual Mapping by a robot rover. In *Pmc. IJCAI'79*, pages 598-600, 1079.

[12] Larry Matthies and Masatoshi Okutomi. A bayesian foundation for active stereo vision. In *SPIE, Sensor Fusion II: Human and Machine Straregies*, pagcs 62-74. Nov. 1989.

[13] M. Yachida. Y. Kitamura. and M. Kimachi. Trinocularvision: New approach for correspondence problem. In *Pmc. ICPR*. pages 1041-1044, 1986.

[14] Victor J. Milenkovic and Takeo Kanade. Trinocular vision using photometric and edge orientation constraints. In *Proceedings of the Image Understanding Workshop*, pagcs 163-175. Miami Beach. Florida. Dec. 1985.

[15] N. Ayache and F. Lustman. Fast and reliable passive trinocular stereo vision. In *Proc. ICCV'87*, pagcs 422-426, 1987.

[16] Roger Y. Tsai. Multiframe image point matching and 3-d surface reconstraction. IEEE *Transacrions on Pattern Analysis and Machine Intelligence*, PAMI-5(No.2), March 1983.

[17] Masatoshi Okutomi and Takeo Kanade. A locally adaptive window for signal matching. In *Proc. of Int'l Conf. on Compurer Ksion*. Dec. 1990.

Part II: Experiments on Outdoor Scenes

T. Nakahara and T. Kanade

In "Part I: Theory." we explained how multiple stereo pairs with various baselines were used W obtain precise depth estimates without suffering from ambiguity. The algorithm was tested wiih indoor images which were taken under well controlled conditions in the Calibrated Imaging Laboratory.

This algorithm is applied to outdoor scenes including variable lighting conditions and large depth range. While Okutomi and Kanade used stereo pairs acquired by moving a camera horizontally, we use stereo pairs raken by moving a camera in borh horizontal and vertical orientations. Taking stereo images with two orthogonal baseline orientations removes ambiguity and increases precision without suffering from rhe orientation of the features in a scene. And we also show that the shapes of the sum of squared-difference (SSD) values near the estimate may indicate the reliability of the match, and suggest a method to classify matches into various types, such as good matches and mismatchrs with occlusion or sparsefeatures.

1. Horizontal baselines Experiment

The experimental setup for acquiring stereo pairs is illustrated in fig. 1. The images are acquired by moving a camera horizontally. The distance between adjacent camera positions is constant. Table 1 describes the image acquisition parameters. Typically, the distance from the camera to the nearest object is 19 m and the baseline length ranges from 19.05 mm for the closest camera pair to 1 14.1 mm for the farthest.

As illustrated in fig. 2. first the input images an: preprocessed with Laplacian of Gaussian (LOG) filter to reduce photometric distortion. A 5x5 window is used for Gaussian and a 3x3 window is used for Laplacian. Then the multiple-baseline stereo is used to compute the inverse depth with a 9x9 window for SSD computation. Typically, the number of the stereo pairs is 6. the image size is 240x256, and the total disparity range is 9 pixels. as summarized in table 2.

1.1. Results

We experiment with three data sets. "Shrubbery," "Parking mctcrs," and "Sand" for the horizontal expenmenu.

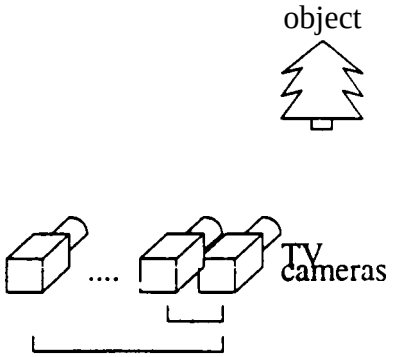


Fig. 1 Setup for horizontal baselines

1.1.1. Shrubbery

Fig. 4 shows the "Shruhhcry" data set which consists of six stereo pairs. The maximum disparity between the adjacent images is around two pixels. Fig. 5 is a LOG preprocessing result of one of the images. Fig. 6 is the isometric plot of the resultant depth map. We observe that the shrubberies at the left and in the center are well separated, and the dcpth jump around the sign board and the top of the signpost are clearly distinct from the wall. We can see a round shrubbery at the right and some pebbles on the road. Some mismatches are observed at the curb. because the features in this area are almost parallel to the epipolar line.

1.12. Parking meters

This data set includes seven stereo pairs. Fig. 8 is the isometric plot of the depth map. The following portions in the scene are well estimated: the three parking meters in front of the shrubberies. the side view of the sign board which is between the second and the third parking meters, and the large dcpth gap between the front and the back building. There are some mismatchcs at the back door of the car because of sparse features in this part.

1.1.3. Sand

This scene contains natural rough surfaces likc sand and a rock as shown in fig. 9. Five stereo pairs are used for this data set. Fig. 10 is the isometric plot of the depth map. We observe that the two rocks and the sand are well estimated. Many mismatches. however, occur at the border between the black wall and the white curtain. The features in this portion are parallel to the epipolar line and are low in density.

12. Shapes of SSD and SSSD Curves

In this section we show that the shapes of the SSDs-in-inverse-depth may indicate the reliability of a match and suggest the cause of a mismatch. For this purpose. we examine the shapes of the SSD and the sum of the SSD (SSSD) in three typical cases: a good match, a mismatch with occlusion. and a mismatch with sparse features.

First. we examine the shapes of the SSD and the SSSD for a point whose depth is precisely and accurately estimated. such as a point ¿ on the sand in fig. 9. Fig. 11 plots 12 curves of individual SSDs and the resultant SSSD for this point. We observe that the minimum of the SSD of each baseline takes place at the same position and the curvature of the SSD near the minimum of the SSSD becomes sharper as the baseline becomes longer. The SSSD exhibits a unique and clear minimum at the correct matching position.

Let us approximate individual SSD's curves by a quadratic equation near the minimum position. From equations (22) - (29) in "Part I: Theory," we expect the following:

- The inverse depth at which the SSD values take the minimum is expected to be the same over the various baselines.
- The curvature is proportional to the square of the hascline length.
- Variance of differences between the inverse dcpth at the minimum position of each SSD and the final estimated inverse depth is inversely proportional to the square of the baseline length.

Fig. 12 (a), (b). and (c) show the above-mcndoncd theoretically expected values and experimental measurements for the case of a good match shown in fig. 11. The measurements coincide well with the theoretical values.

Second, we look into the occlusion case, such as a point ¡ at the right of the first parking meter head in fig. 7. The correspondence points exist in shorter baselines. As the baseline becomes longer, occlusion. however, occurs and matching is nor possible. The SSD and the SSSD for the point ¡ are shown in fig. 13 (a). The inverse depth at the minimum of the SSD of each baseline gradually shifts from the true position to a false position. The SSSD does not show a clear minimum. As shown in fig.13 (b), (c), and (d), the theoretically expected values and the measurements coincide where the baselines are short but differ greatly where the haselines are long.

The third case is a point with sparse features like a point ¬ at the black wall in fig. 9. As shown in fig. 13(e), the SSD curve of each hascline is almost flat over the inverse depth range with no obvious minimum. Consequently the SSSD docs not have the minimum.

Another observation for the part of a depth map wilh mismatch or noisy measurements is the problem of the orientation of the features in a scene. We can not obtain good depth estimates near the curb portion in "Shrubbery" or the border between the black

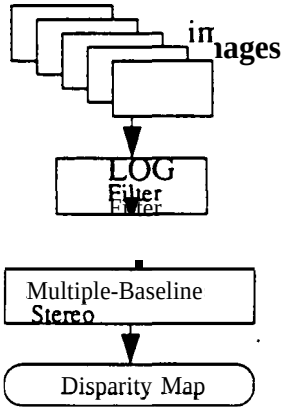


Fig. 2 Procedure

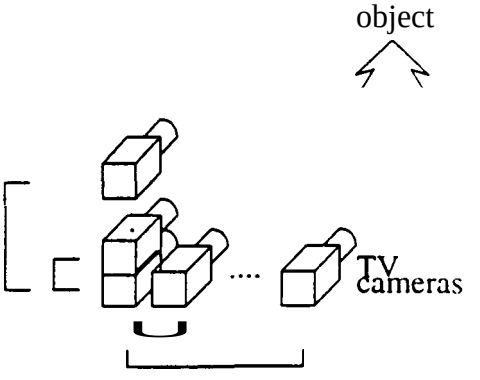


Fig. 3 Setup for horizontal and vertical baselines

wall and the curtain in "Sand." because the image contains only horizontal features. The solution is to use additional stereo image pairs taken by cameras aligned in vertical direction. Combining the information of the vertical baseline with the information of the horizontal baseline is straightforward in the multiple-baseline algorithm. because this algorithm simply adds the SSD-in-inverse-depth instead of the disparity.

Next section. we demonstrate the effectiveness of using both horizontal and vertical baselines.

2. Horizontal and vertical baselines Experiment

Fig. 3 illustrates the experimental setup. The procedure is the Same as the one in the horizontal baselines experiment. except images are taken by moving a camera horizontally and vcnically. The acquisition parameters are shown in the last three rows ("Shrubbery2," "Comer." and "Guide") of table I. The baseline length ranges from 20 mm for the closest camera pair to 60 mm for the farthest, which is somewhat shorter than the horizontal baselines case.

As shown in the last three rows in table 2. the number of the stereo pairs is 3 for each baseline and the total disparity range is 6 pixels.

2.1. Results

2.1.1. "Corner" data set

Fig. 14 shows the data set. togcthcr with an illustration of the arrangement of the camera and the objects.

Fig. 16 is the isometric plot of the dcpth map using three stcrco pairs for each baseline orientation (six pairs in total). We observe that the building wall. especially the slanting part of the wall at the right is well estimated. The curb is separated from the shrubberies in the back and the road in the front. We can see the distances between the curb and the shrubberies.

For comparison, a depth map is computed using six stcrco pairs in only horizontal orientation. Fig. 15 is the result. Many mismatches are observed at the wall and the curb. because the main features of these portions are horizontal.

2.1.2. SSD and SSSD in inverse depth

We examine the SSD and the SSSD of a point. such as a point on the wall or at the curb. Depth estimate of the point is corrcct using both horizontal and vertical baselines. though the estimate is incorrect using only horizontal baselines. Fig. 17 (a) and (b) show the SSD and the SSSD of the points ¿ and ¡ in fig. 16 rcspcctivcly. Though the SSDs of the horizontal baselines do not show the clear minimum. the SSDs of the vertical baselines having a unique minimum at the correct point contribute to the determination of the correct minimum position of the total SSSD.

3. Comments

Parallelism

The computation of this algorithm is simple and local. which is suited for implementation on a massively parallel machine. We are implementing this algorithm on a MasPar, a Single Instruction Multiple Data (SIMD) machine with 4096 processors. At this moment, the processing time on MasPar is 0.9 second for producing a depth map with the 240x256 image size and the disparity range of 10 pixels. while it takes 51 seconds to do the same for SUN 4/40(16 MIPS).

It is possible to implement this algorithm by a dedicated hardwan: or cvcn on a chip for a real-time depth sensor.

Classification of depth measurements

We showcd that the shapes of the SSD in-inverse-dcpth near the minimum of the SSSD may be useful to estimate the cause of mismatches like the occlusion and the sparse features cases. We expect that we can similarly analyze mismatchcs caused by terminal edge or a highlight.

Acknowledgment

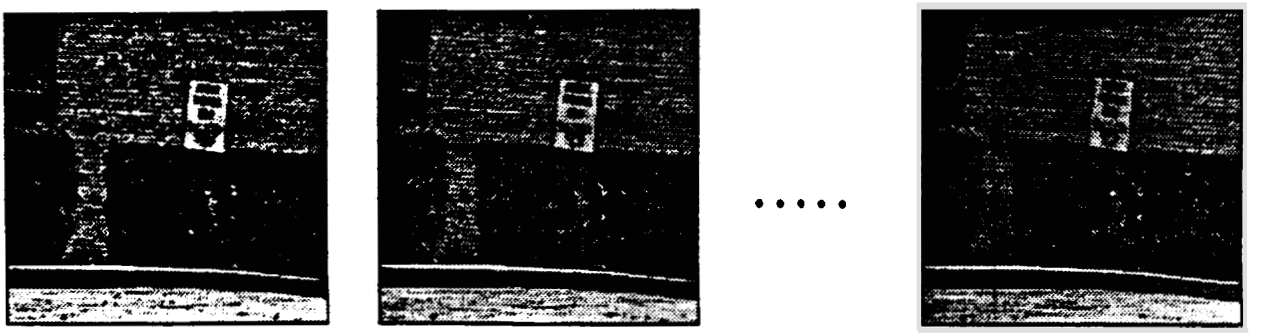
The authors would like to thank Kenichi Arakawa, Eric Krotkov, and Hans Thomas for assisting in imagc acquisition and reading the manuscript.

Table 2 Image processing

Name	Number of stereo pair	Image size	Disparity range
Town	3	256x240	4 - 14
Coal	5	256x240	30 - 40
Shrubbery	6	240x256	4 - 13
Parking meters	7	240x256	1 - 15
Sand	5	240x256	1 - 6
Shrubbery2	H : 3 v : 3	240x256	1 - 7
Corner	H : 3 v : 3	240x256	1 - ?
Guide	H : 3 v : ■	240x256	0 - x

Table 1 Image acquisition

Name	Distance		Baseline length		TV camera	Focal length
	the nearest	the farthest	unit	longest		
Town	0.51m	1 02m	1.27mm	11 43mm	——	——
coal	——	——	7 62mm	38 10mm	——	——
Shrubbery	19m	28m	19 05mm	114 10mm	SONY XC57	50mm
Parking meters	12m	34m	10 16mm	71 12mm	SONY XC57	50mm
Sand	6m	10m	2 54mm	12 70mm	SONY SSC-D7	50mm
Shrubbery2	19m	28m	20 0mm	60.00mm	SONY xc57	50mm
Comer	19m	28m	20 0mm	60 00mm	SONY XC57	50mm
Guide	16m	90m	20 0mm	60.00mm	SONY XC57	50mm



(a) 1 st (left most) (b) 2 nd (c) 7 th (right most)

Fig. 4 "Shrubbery" data

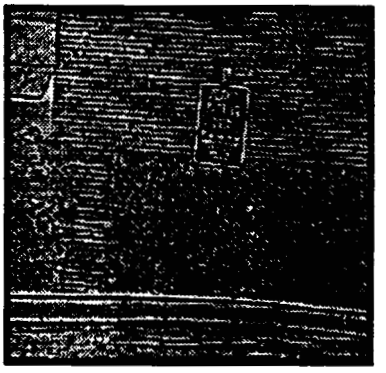


Fig. **5** Laplacian of Gaussian image

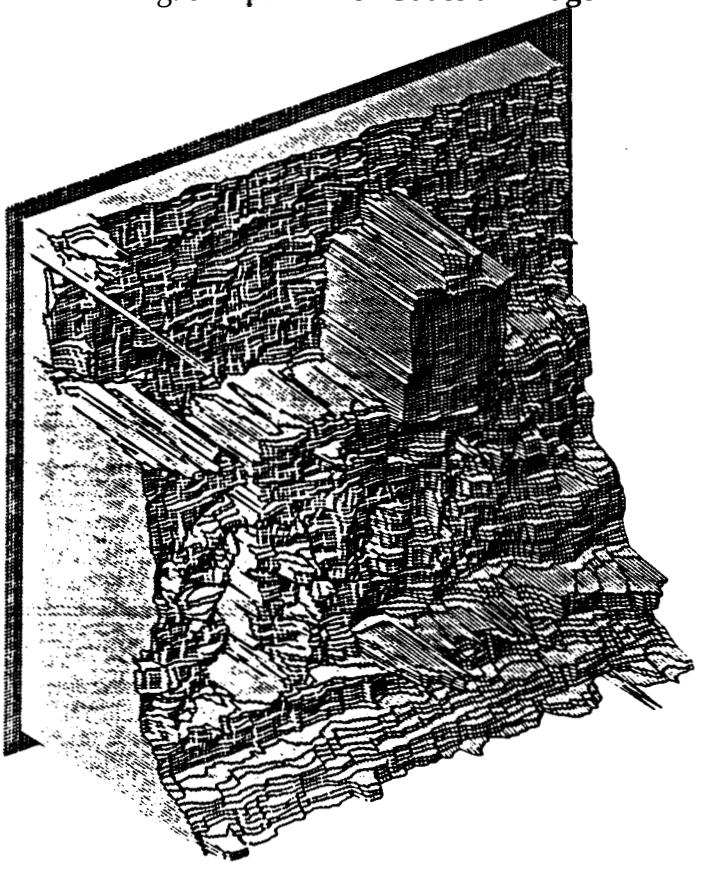


Fig. **6** Isometric plot of depth **("Shrubbery")**

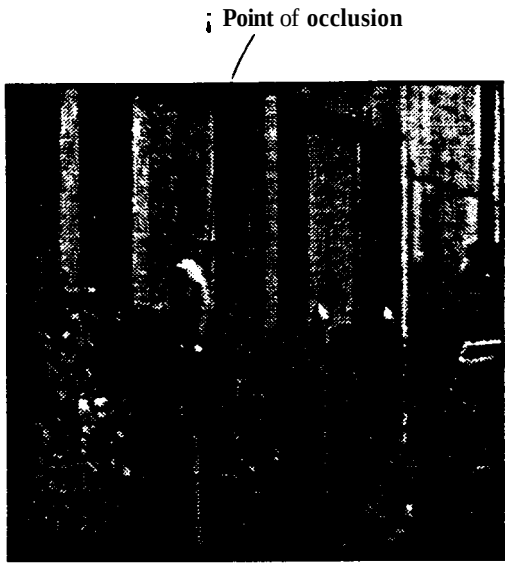


Fig. 7 "Parking meters" data

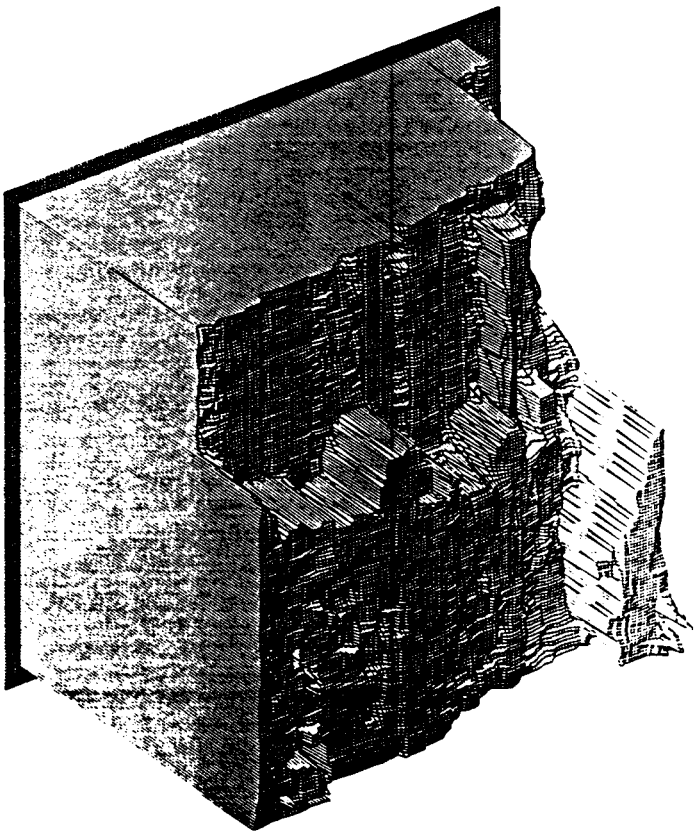


Fig.8 Isometric plot of depth ("Parking meters")

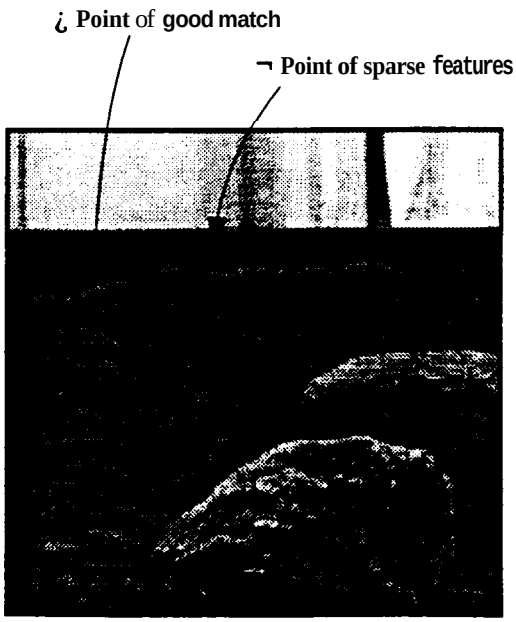


Fig. S "Sand" data

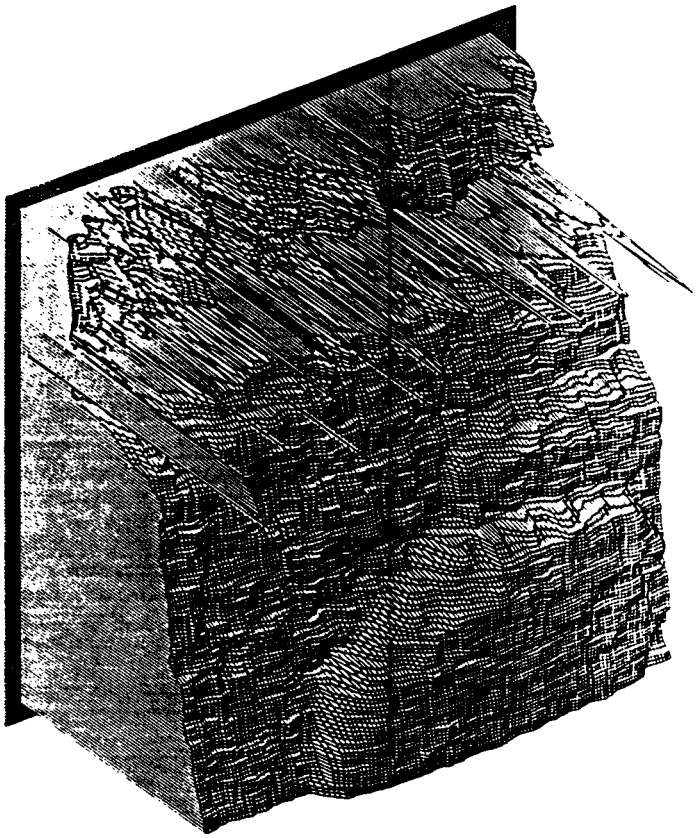
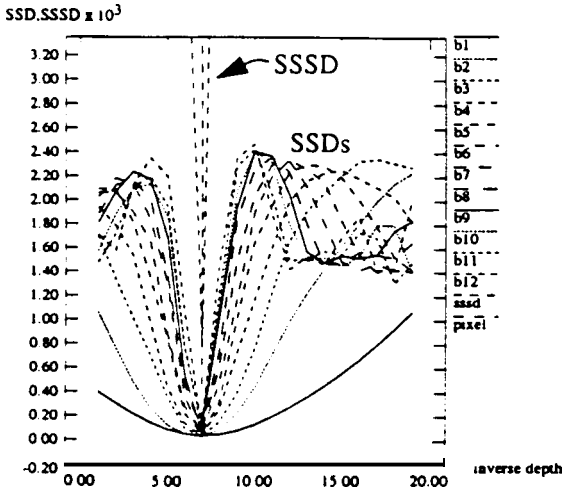
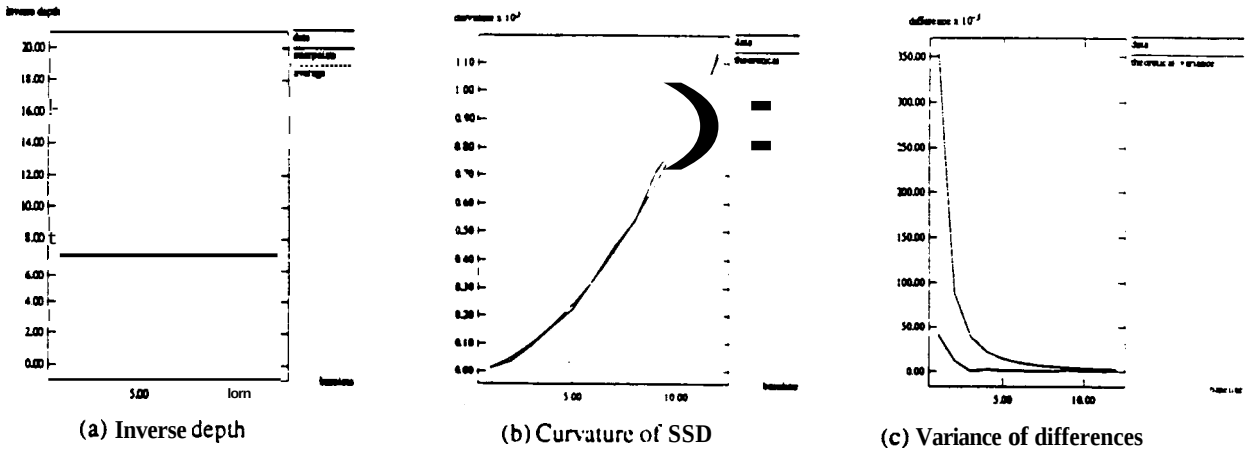


Fig.10 Isometric plot of depth ("Sand")



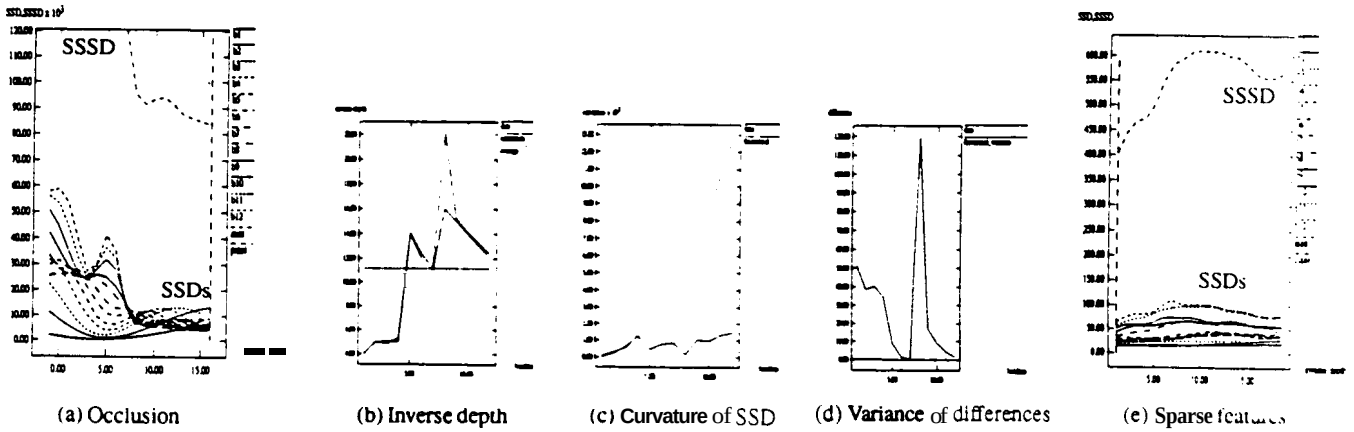
(a) Good match

Fig. 11 **SSD** and **SSSD values** vs. inverse depth



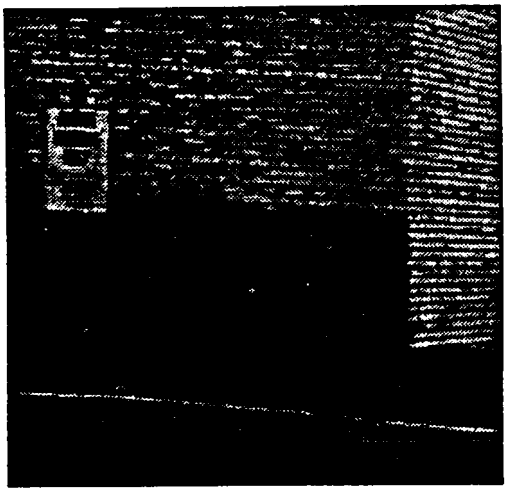
(a) Inverse depth (b) Curvature of SSD (c) Variance of differences

Fig. 12 Inverse depth, variance of differences between the minimum position of each **SSD** and the final esumate. and curvature of SSD trom individual stereo pair near the minimum of **SSSD** vs. baseline (Case of a **good** match)

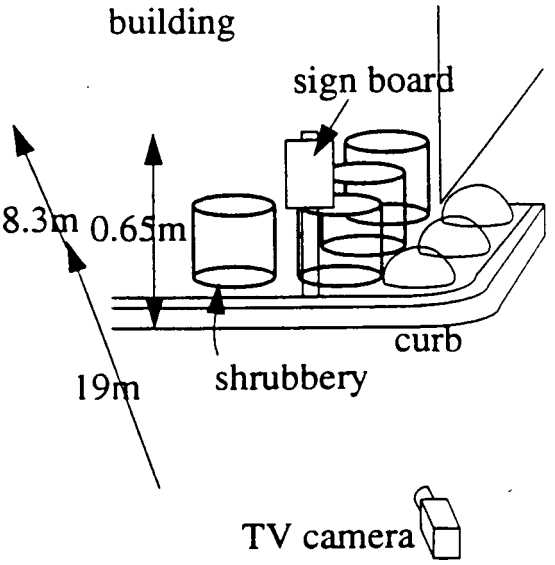


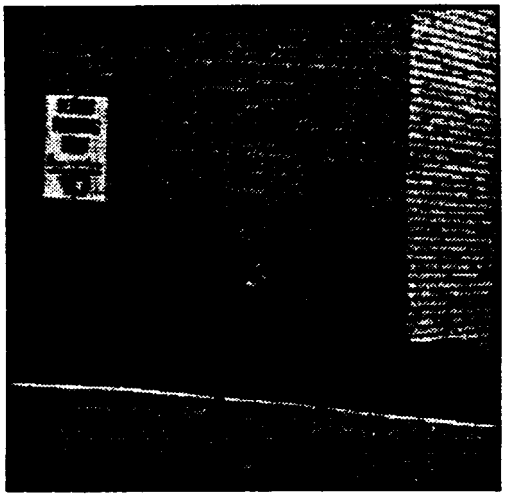
(a) Occlusion (b) Inverse depth (c) Curvature of SSD (d) Variance of differences (e) Sparse features

Fig. 13 **SSD** and **SSSD** values vs. inverse depth. inverse depth. variance of differences between the minimum position of each **SSD** and the final estimate. and curvature of **SSD** from individual stereo pair near the minimum of **SSSD** vs. baseline

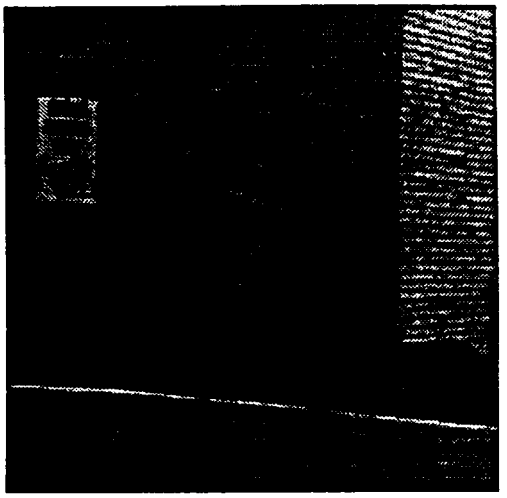


(a) Up most





(b) Left most



(c) Right most

Fig. 14 "Corner" **data**

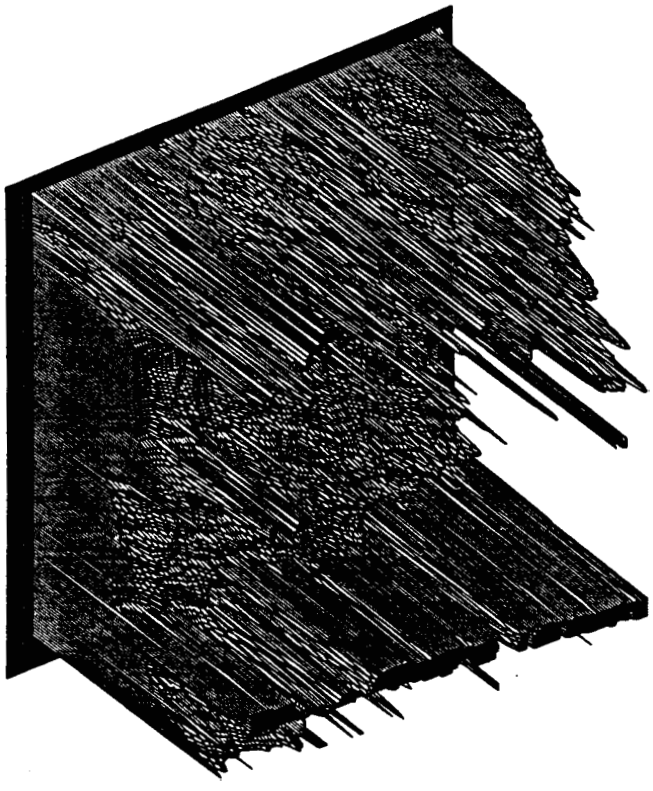


Fig. 15 Isometric plot of depth resulted from horizontal baselines

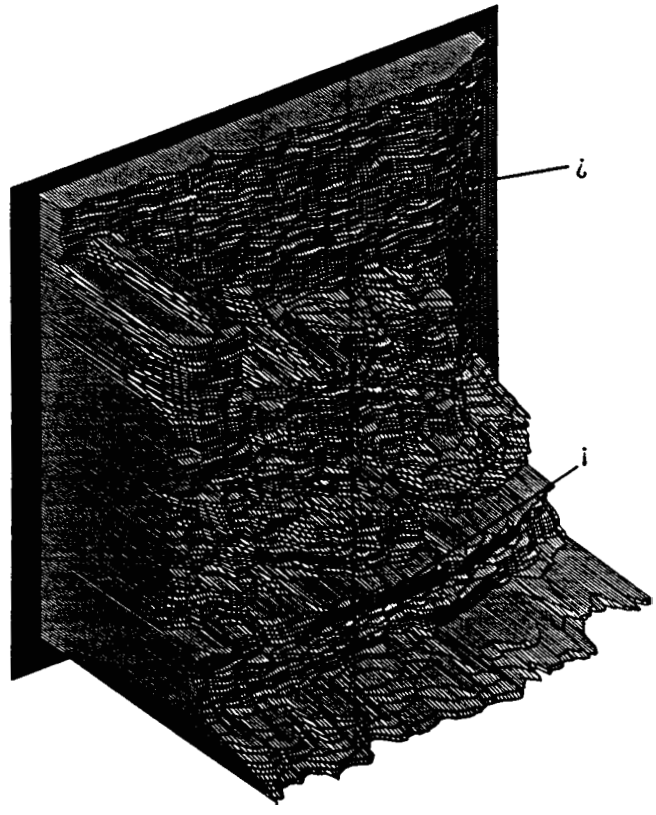
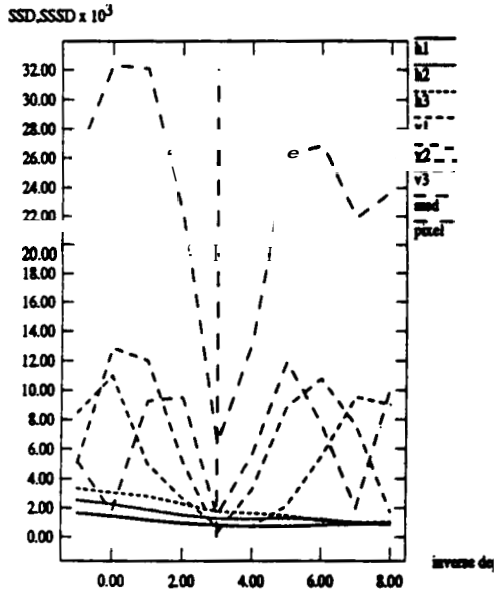
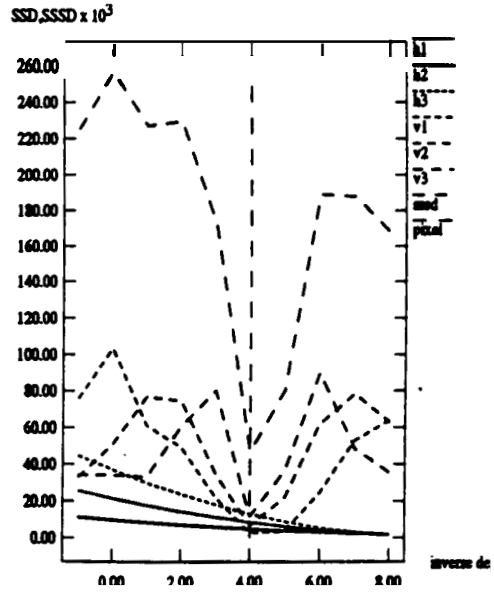


Fig. 16 Isometric plot of depth resulted from horizontal and vertical baselines



(a) On the wall



(b) At the curb

Fig. 17 SSD and SSSD values vs. inverse depth