A Multiple-Baseline Stereo

Masatoshi Okutomi [1] and **Takeo Kanade**

November 28, 1990

CMU-CS-90-189$_2$

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

This research was supported in part by the Defense Advanced Research Projects Agency (DOD), and monitored by the Avionics Laboratory, Air Force Wright Aeronautical Laboratories, Aeronautical Systems Division (AFSC), Wright-Patterson AFB, 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 authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or of the U.S. Government.

---

[1] This research was performed while the first author was with Carnegie Mellon University. His permanent address is Information Systems Research Center, Canon Inc., 890-12 Kashimada, Saiwai-ku, Kawasaki, 211, Japan

Keywords: vision, image understanding, stereo

Abstract

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

In stereo processing, a short baseline means that the estimated depth will be less precise due to narrow triangulation. For more precise depth estimation, a longer baseline is desired. With a longer baseline, however, a larger disparity range must be searched to find a match. As a result, matching 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 by computing the sum of squared-difference (SSD) values. The SSD functions for individual stereo pairs are represented with respect to the inverse depth (rather than the disparity, as is usually done), and then are simply added to produce the sum of SSDs. This resulting function is called the SSSD-in-inverse-depth. We show that the SSSD-in-inverse-depth 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 matches and increase precision without any search or sequential filtering.

This paper first defines a stereo algorithm based on the SSSD-in-inverse-depth and presents a mathematical 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 distance between the corresponding points of left and right images. The disparity d is related to the depth z by

$$d = BF\frac{1}{z} \tag{1}$$

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

This equation indicates that for the same depth 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 depth 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 searched, 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 [MP79][Gri85]. Matching is done at a low resolution to reduce false matches and then the result is used to limit the search range of matching at a high 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 part 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 [BBM87, Yam88, MSK89]. A short 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 [BBM87, Yam88] or by a sequential Kalman filtering technique [MSK89, Hee89].

A 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. Matching, however, is performed in a simple way. The sum of square-difference (SSD) values are computed for each pair of stereo images. The SSD values are represented with respect to the inverse depth $\frac{1}{z}$ (rather than the disparity d, as is usually done). Those SSD functions from all the stereo pairs are simply added together to produce the sum of SSDs, which we call SSSD-in-inverse-depth. We show that the SSSD-in-inverse-depth 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 technique is that we can eliminate false matches and increase precision without any search or sequential filtering.

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

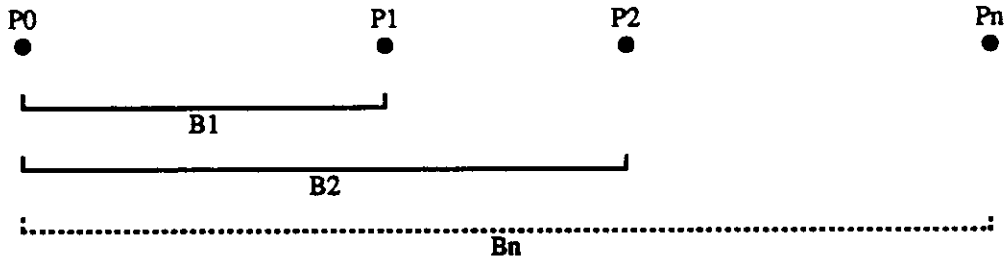


Figure 1: Camera positions for stereo

2 Mathematical Analysis

The essence of stereo matching is, given a point in one image, to find the most similar point in another image. 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 for matching. In this section, we define the sum of SSD with respect to the inverse depth (SSSD-in-inverse-depth) 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 camera positions $P_0, P_1, \ldots, P_n$ 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 depth is z. Its disparity $d_{r(i)}$ for the image pair taken from P_0 and P_i is

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

The image intensity functions $f_0(x)$ and $f_i(x)$ near the matching positions for Z can be expressed 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 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:

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

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 Depth

Now, let us introduce the *inverse depth* 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 depth, respectively. Substituting these into equation (5), we have the SSD with respect to the inverse depth,

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

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

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

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

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

Its expected value is

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

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

$$E[e_{\zeta(i)}(x, \zeta_r)] = E[e_{\zeta(i)}(x, \zeta_r + \frac{a}{B_i F})] = 2N_w \sigma_n^2. \tag{16}$$

We still have an ambiguity; a minimum is expected at a false inverse depth $\zeta_f = \zeta_r + \frac{a}{B_i F}$. However, an important point to be observed here is that this minimum for the false inverse depth ζ_f changes

its position as the baseline B_i changes, while the minimum for the correct inverse depth ζ_r does not. This is the property that the new evaluation function, the SSSD-in-inverse-depth (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)

$$\begin{aligned} E[e_{\zeta(12)}(x_0, \zeta)] &= \sum_{j \in W} (f(x_0 + j) - f(x_0 + B_1 F(\zeta - \zeta_r) + j))^2 \\ &+ \sum_{j \in W} (f(x_0 + j) - f(x_0 + B_2 F(\zeta - \zeta_r) + j))^2 + 4N_w \sigma_n^2 \end{aligned} \tag{17}$$

We can prove that

$$E[e_{\zeta(12)}(x_0, \zeta)] > 4N_w \sigma_n^2 = E[e_{\zeta(12)}(x_0, \zeta_r)] \quad \text{for } \zeta \neq \zeta_r. \tag{18}$$

(refer to appendix A) In words, $e_{\zeta(12)}(x_0, \zeta)$ 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.

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

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

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 $e_{d(1)}(x_0, d(1))$ 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 depth 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 $B_1 F = 1$). Note that the minima at the correct inverse depth ($\zeta = 5$) does not move, while the minima for the false match changes its position as the baseline changes. When the two functions are added to produce the SSSD-in-inverse-depth, its expected values $E[e_{\zeta(12)}]$ are as shown in figure 2 (f). We can see that the ambiguity has been reduced because the SSSD-in-inverse-depth has a smaller value at the correct match position than at the false match.

2.4 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.

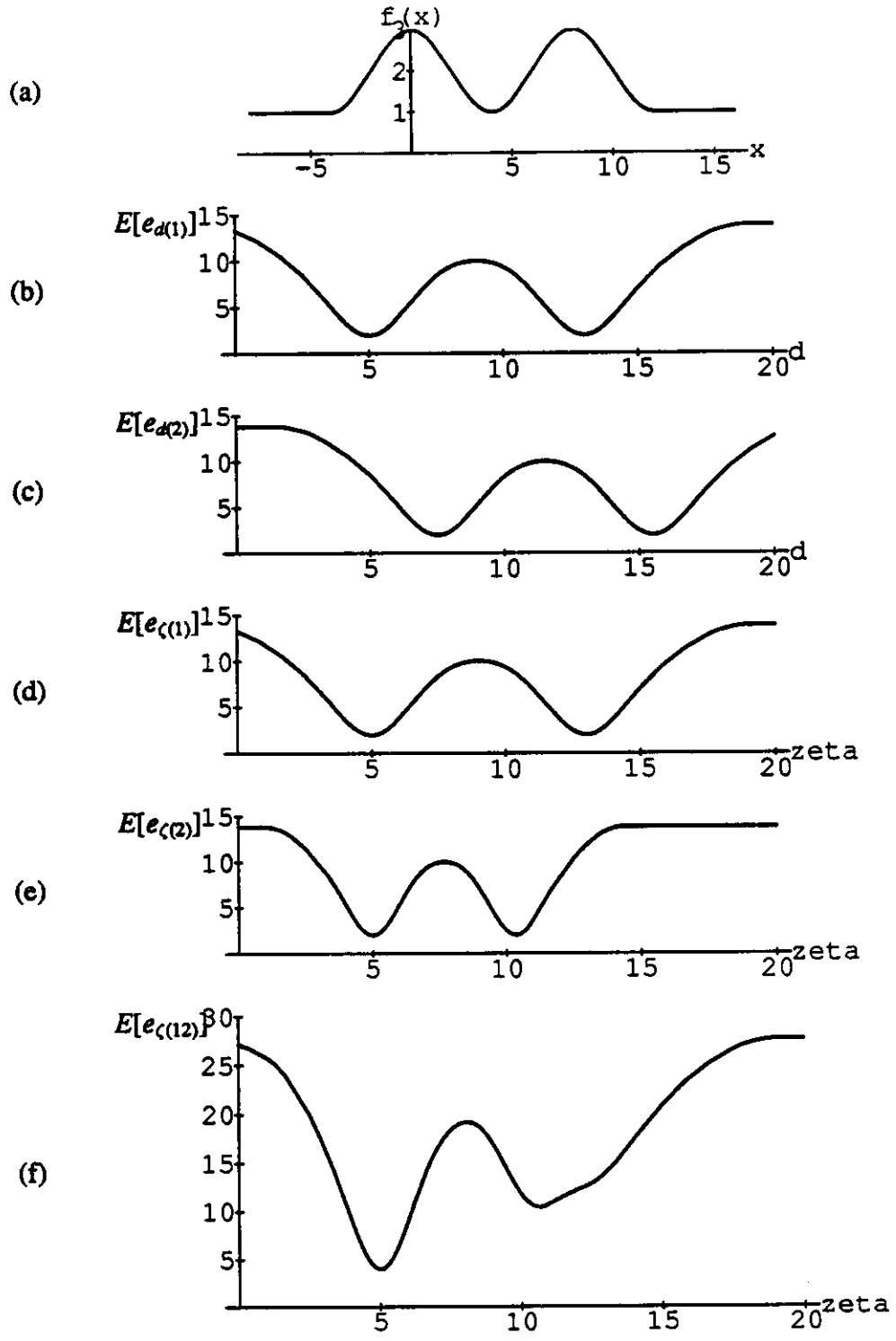


Figure 2: Expected values of evaluation functions: (a) Underlying function; (b) $E[e_{d(1)}]$; (c) $E[e_{d(2)}]$; (d) $E[e_{\zeta(1)}]$; (e) $E[e_{\zeta(2)}]$; (f) $E[e_{\zeta(12)}]$

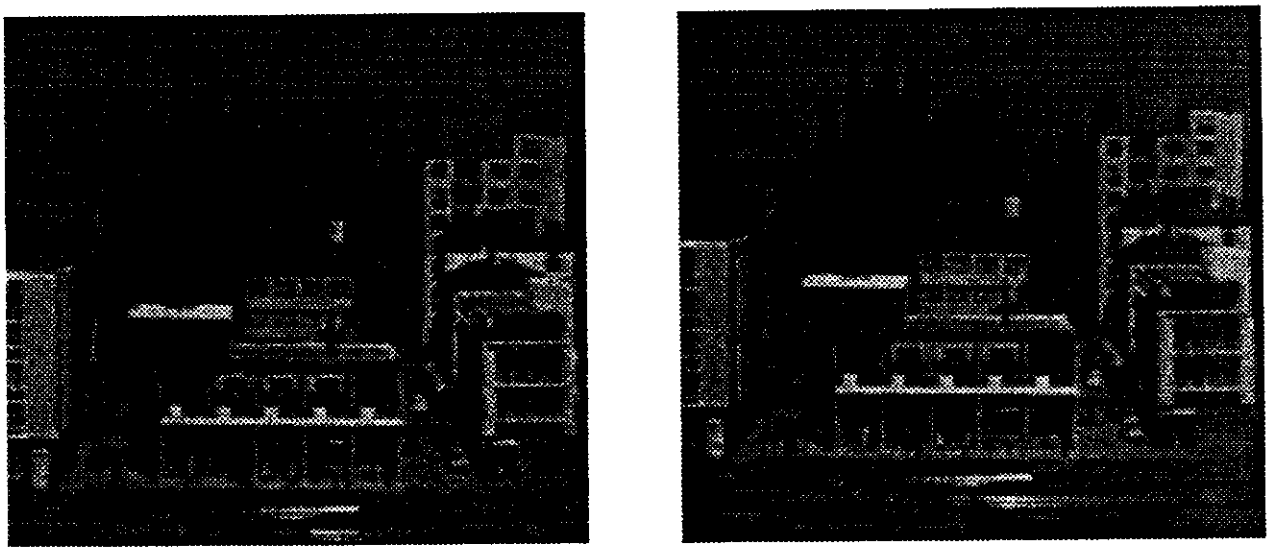
Let $f(x)$ be a periodic function with period T. Then, each $e_{\zeta(i)}(x, \zeta)$ is expected to be a periodic function of ζ with the period $\frac{T}{B_i F}$ (for the proof, refer to appendix B). This means that there will be multiple minima of $e_{\zeta(i)}(x, \zeta)$ (i.e., ambiguity in matching) 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, but its period T_{12} is increased to

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

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 stereo pairs. Furthermore, it can be controlled by choosing the baselines 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 stereo may suffer from ambiguity.

We illustrate this by using real stereo images. Figure 3(a) shows an image of a sample scene. At the top of the scene there is a grid board whose intensity function is nearly periodic. We took ten images of this scene by shifting the camera vertically 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. The SSD values $e_{\zeta(i)}(x, \zeta)$ over 5-by-5-pixel windows are plotted for various baseline stereo pairs in figure 5. The horizontal axis of all the plots is the inverse depth, normalized such that $8bF = 1$. Figure 5 illustrates the trade-off between precision and ambiguity in terms of baselines. That is, for a shorter baseline, there are fewer minima (i.e. less ambiguity), but the SSD curve is flatter (i.e. less precise localization). On the other hand, for a longer baseline, there are more minima (i.e. more ambiguity), but the curve near the minimum is sharper; that is. the estimated depth is more precise if we can find the correct one.

Now, let us 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 to ∞ inches in depth. 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-baseline 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.



(a) (b)

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

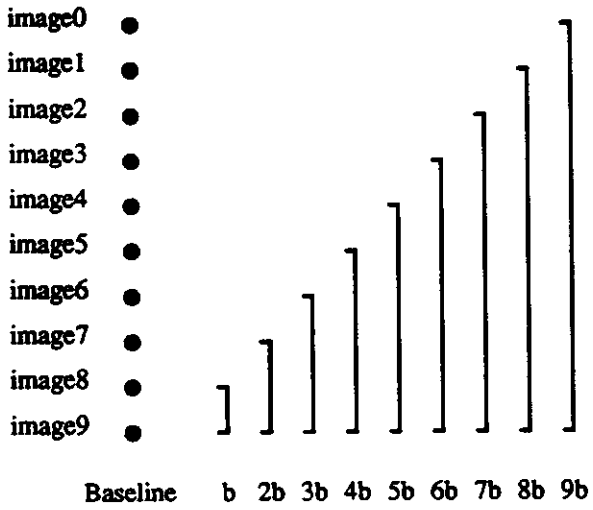


Figure 4: "Town" data set image sequence

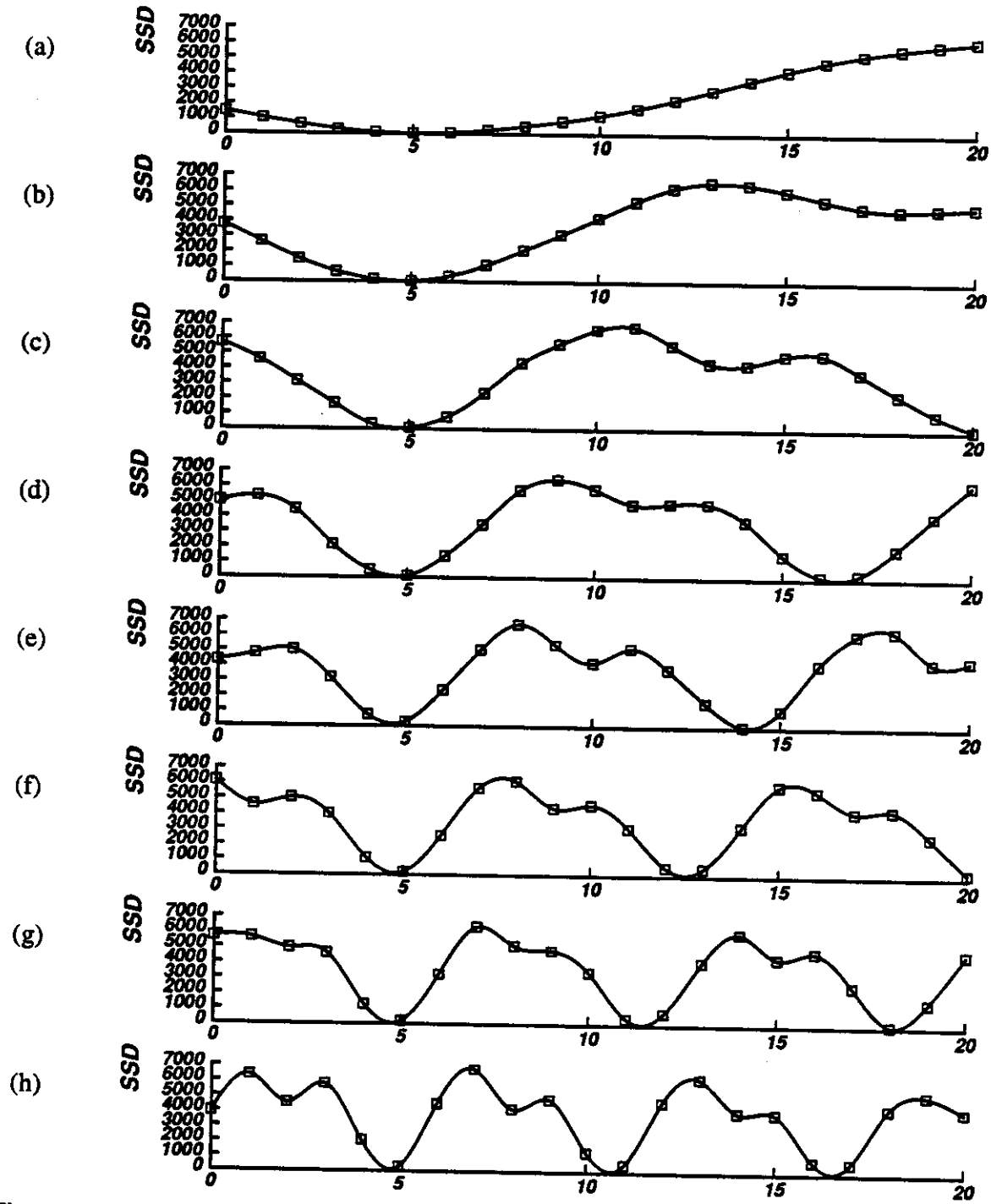


Figure 5: SSD values vs. inverse depth: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

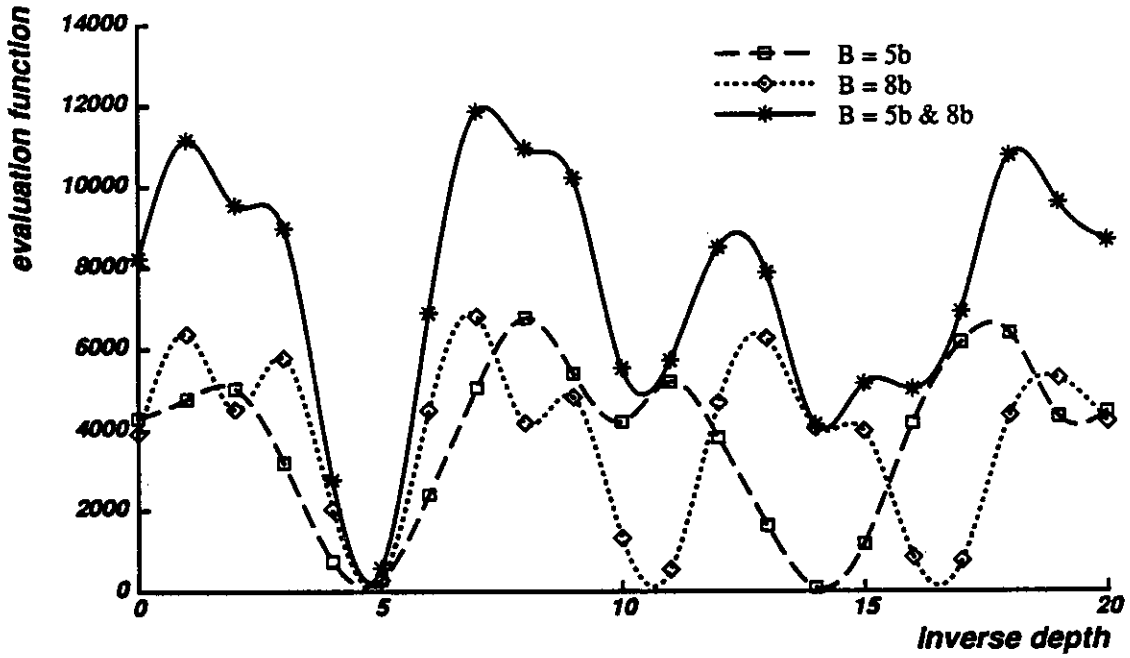


Figure 6: Combining two stereo pairs with different baselines

So far, we have considered using only two stereo 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 $E[e_{C(12\cdots n)}(x,\zeta)]$ becomes

$$T_{12,\ldots,n} = LCM\left(\frac{T}{B_1F}, \frac{T}{B_2F}, \ldots, \frac{T}{B_nF}\right) \tag{21}$$

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 stereo 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-depth shows the minimum at the correct position more clearly as more stereo pairs are used.

2.5 Precision

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

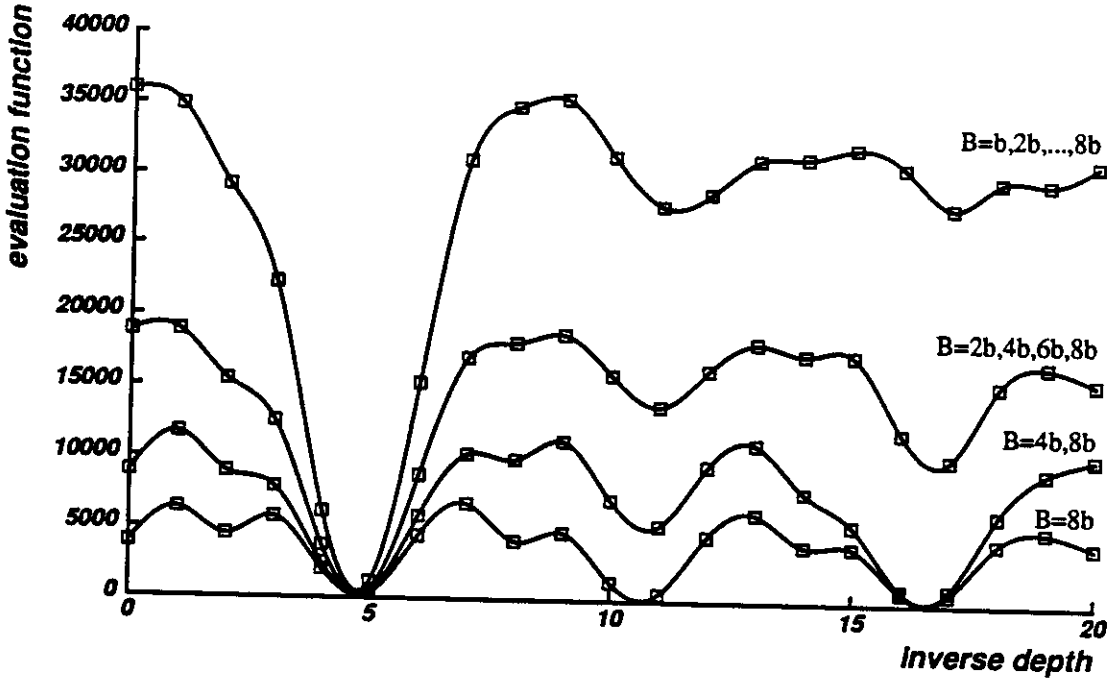


Figure 7: Combining multiple baseline stereo pairs

By using the Taylor expansion about ζ_r, we obtain

$$f(x + B_i F(\zeta - \zeta_r) + j) \approx f(x + j) + B_i F(\zeta - \zeta_r) f'(x + j). \tag{22}$$

Substituting this into equation (12), we can approximate $e_{\zeta(i)}(x, \zeta)$ near ζ_r by a quadratic form of ζ:

$$\begin{aligned} e_{\zeta(i)}(x, \zeta) &\approx \sum_{j \in W} (-B_i F(\zeta - \zeta_r) f'(x + j) + n_0(x + j) - n_i(x + B_i F \zeta + j))^2 \\ &= B_i^2 F^2 a(x)(\zeta - \zeta_r)^2 + 2B_i F(b_i(x) - b_0(x))(\zeta - \zeta_r) + c_i(x), \end{aligned} \tag{23}$$

where

$$a(x) = \sum_{j \in W} (f'(x + j))^2 \tag{24}$$

$$b_k(x) = \sum_{j \in W} f'(x + j) n_k(x + j) \tag{25}$$

$$c_i(x) = \sum_{j \in W} (n_i(x + j) - n_0(x + j))^2. \tag{26}$$

The estimated inverse depth $\hat{\zeta}_r$ is the value ζ that makes equation (23) minimum;

$$\hat{\zeta}_r = \zeta_r - \frac{b_i(x) - b_0(x)}{B_i F a(x)} \tag{27}$$

As shown in (13), the expected value of the estimate $\hat{\zeta}_r$ is the correct value ζ_r, but it varies due to the noise. Following Appendix A in [MSK89], the variance of this estimate is given by

$$Var_{(i)}(\hat{\zeta}_r) = \frac{2\sigma_n^2}{B_i^2 F^2 a(x)}. \tag{28}$$

Basically, for the same amount of 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

$$e_{\zeta(12\cdots n)}(x, \zeta) \approx \left(\sum_{i=1}^{n} B_i^2\right) F^2 a(x)(\zeta - \zeta_r)^2 + 2\left(\sum_{i=1}^{n} B_i F(b_i(x) - b_0(x))\right)(\zeta - \zeta_r) + \sum_{i=1}^{n} c_i(x). \tag{29}$$

The variance of the estimated inverse depth $\hat{\zeta}_r$ that minimizes this function is

$$Var_{(12\cdots n)}(\hat{\zeta}_r) = \frac{2\sigma_n^2}{\left(\sum_{i=1}^{n} B_i^2\right) F^2 a(x)}. \tag{30}$$

From equations (28) and (30), we see that

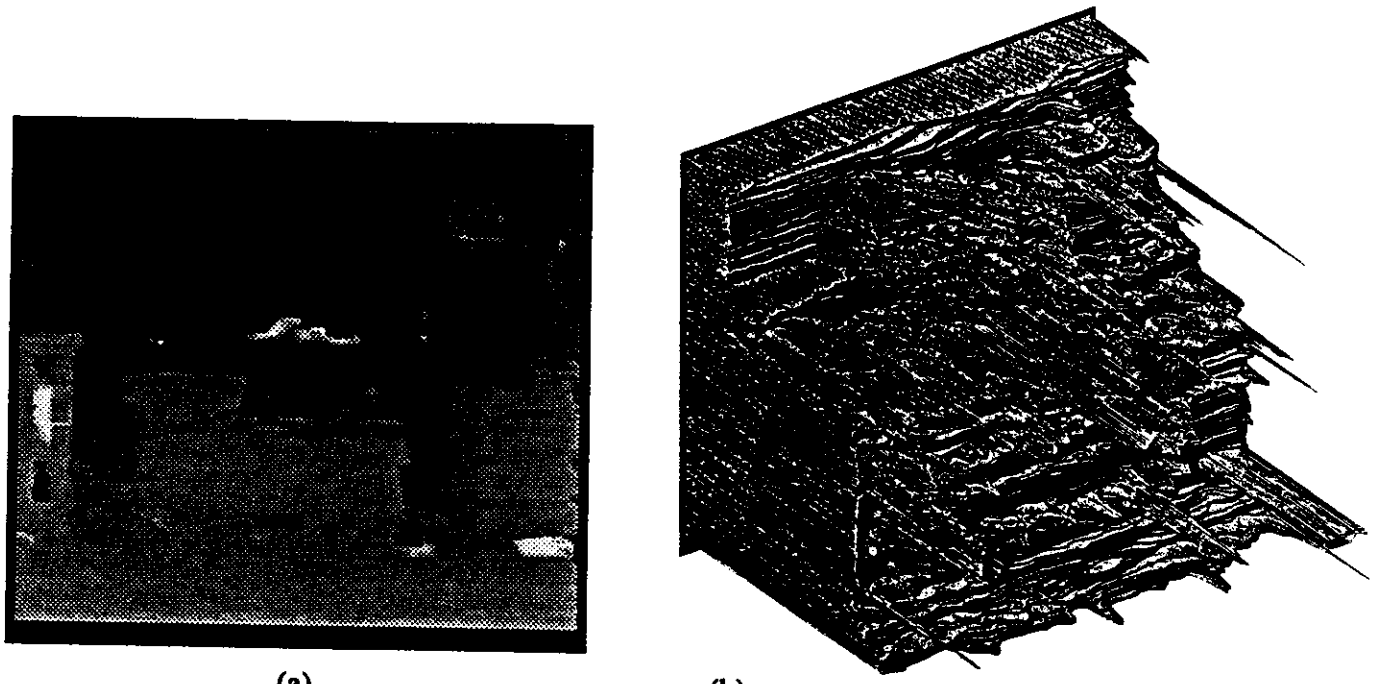
$$\frac{1}{Var_{(12\cdots n)}(\hat{\zeta}_r)} = \sum_{i=1}^{n} \frac{1}{Var_{(i)}(\hat{\zeta}_r)}. \tag{31}$$

The inverse of the variance represents the precision of the estimate. Therefore, equation (31) means that by using the SSSD-in-inverse-depth 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 correct inverse depth becomes sharper as more baselines are used.

3 Experimental Results

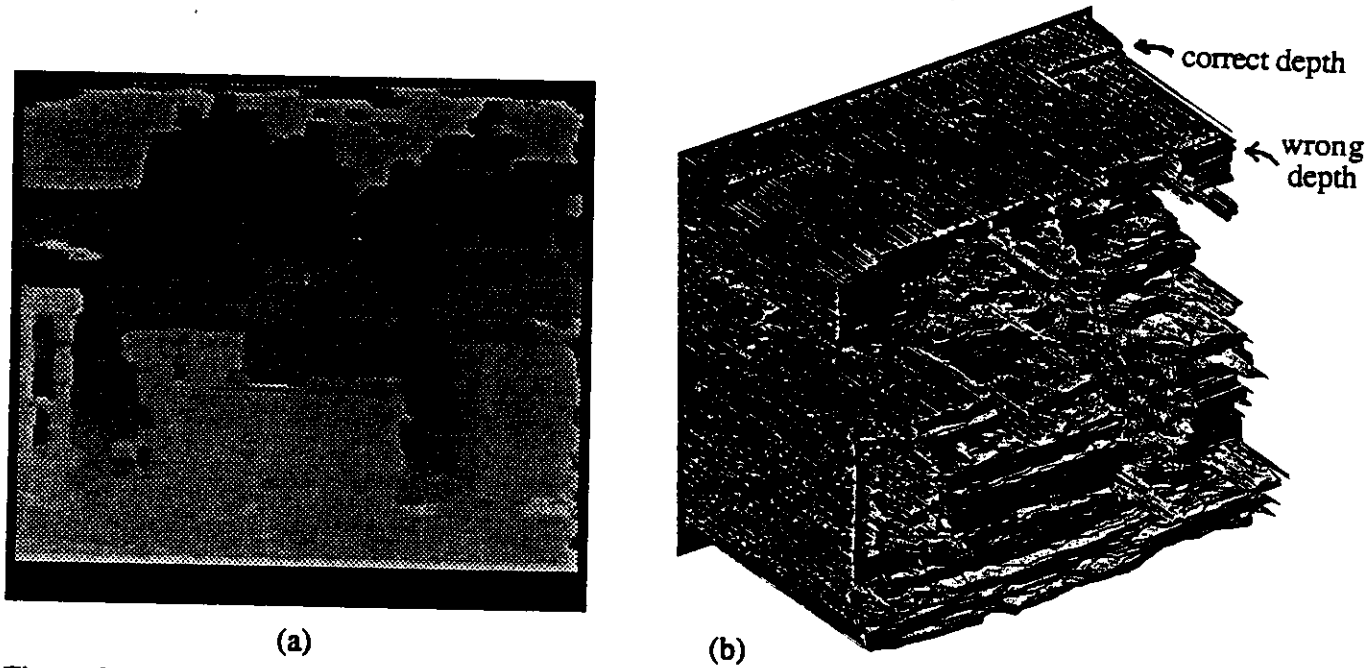
This section presents experimental results of the multiple-baseline stereo based on SSSD-in-inverse-depth with real 2D images. A complete description of the algorithm is included in Appendix C.

The first result is for the "Town" data set that we showed in figure 3. Figures 8 (a) and (b) are the depth 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 depth 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 depth map and its isometric plot obtained by the new algorithm using three different baselines, $3b$, $6b$, and $9b$. For comparison, the corresponding oblique view of the scene is shown in figure 11. We can note



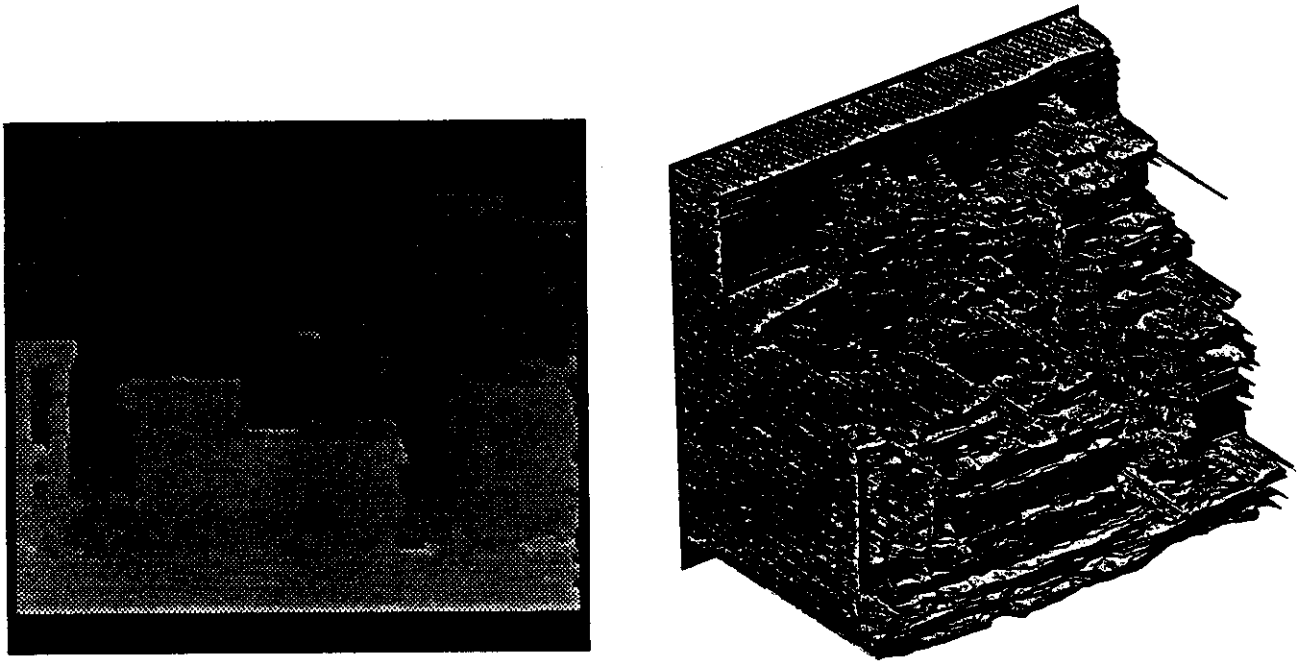
(a) (b)

Figure 8: Result with a short baseline, $B = 3b$: (a) Depth map; (b) Isometric plot of the depth map from the upper left corner. The matching is mostly correct, but very noisy.



(a) (b)

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

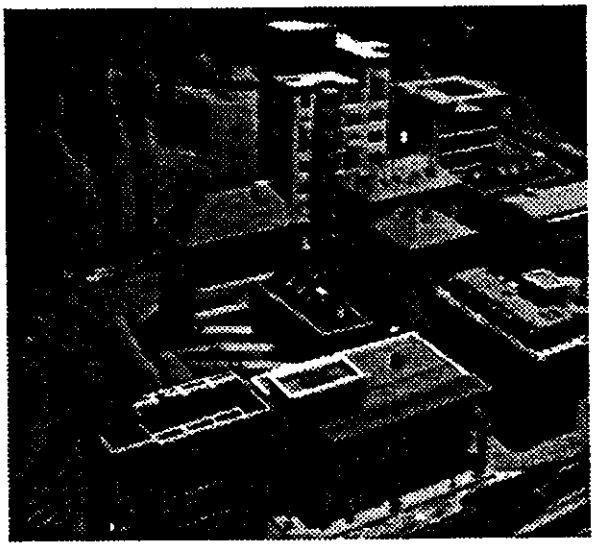
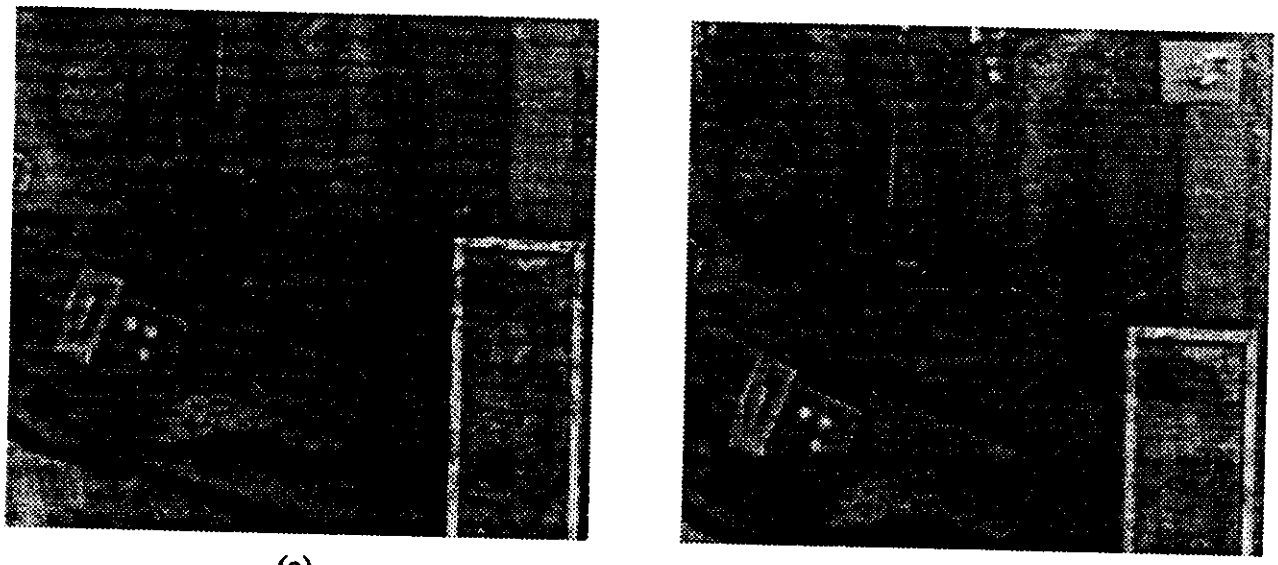


Figure 11: Oblique view



(a) (b)

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

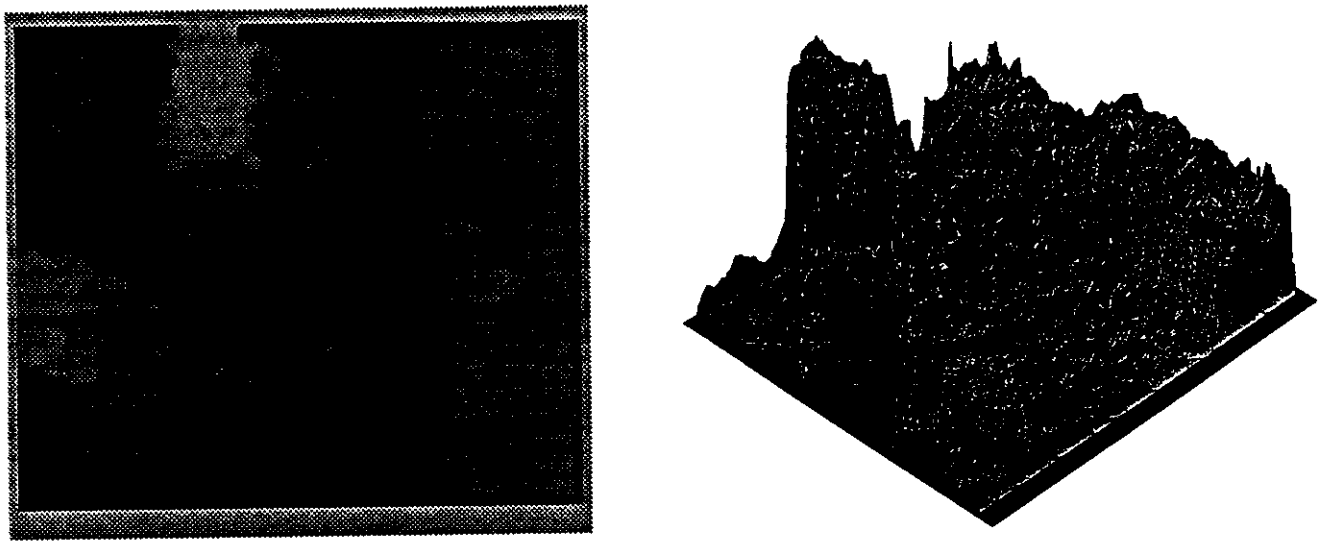
that the computed depth map is less ambiguous *and* more precise than those of the single-baseline stereo.

Figure 12 shows another data set used for our experiment. Figures 13 and 14 compare the depth maps computed from the short 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 depth map obtained with the long baseline due to false matching. In contrast, the result from the multiple-baseline stereo shown in figure 16 demonstrates both the advantage of unambiguous matching with a short baseline and that of precise matching with a long baseline.

Finally, figure 17 shows the depth map obtained by our stereo matching algorithm with a locally adaptive window, which has been presented in [KO90], for which the depth map of figure 16 is used as the initial estimate. The map exhibits both smoothness for flat surfaces and sharpness at depth edges.

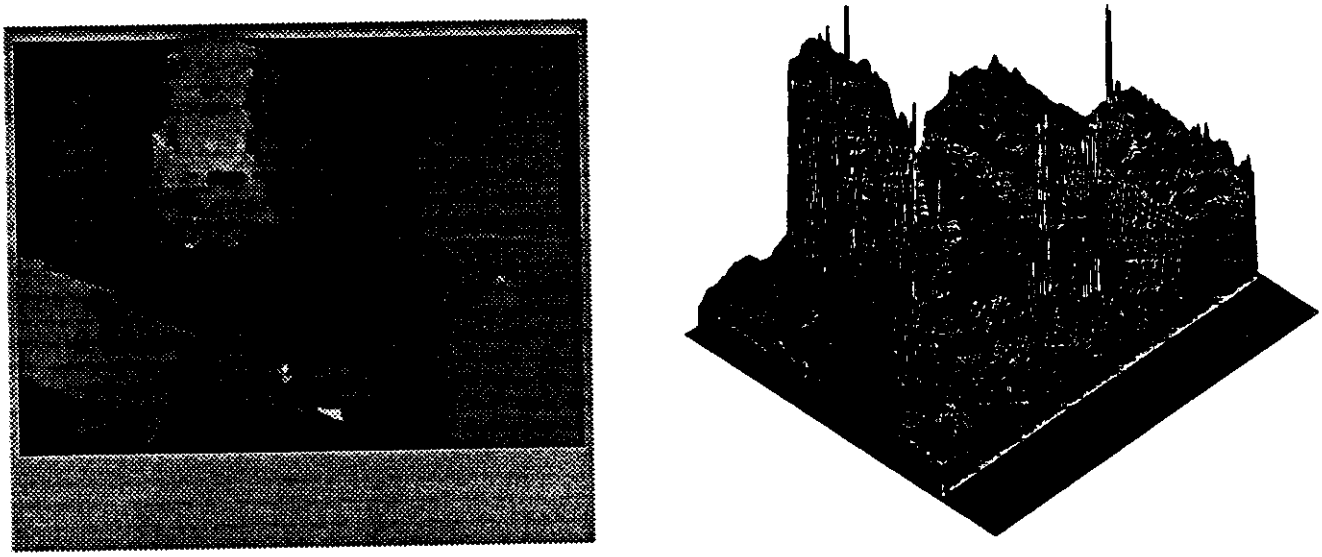
4 Conclusions

In this paper, we have presented a new stereo matching method which uses multiple baseline stereo 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 depth, and add those functions. The resulting



(a) (b)

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



(a) (b)

Figure 14: Result with a long baseline: (a) Depth map; (b) Isometric plot

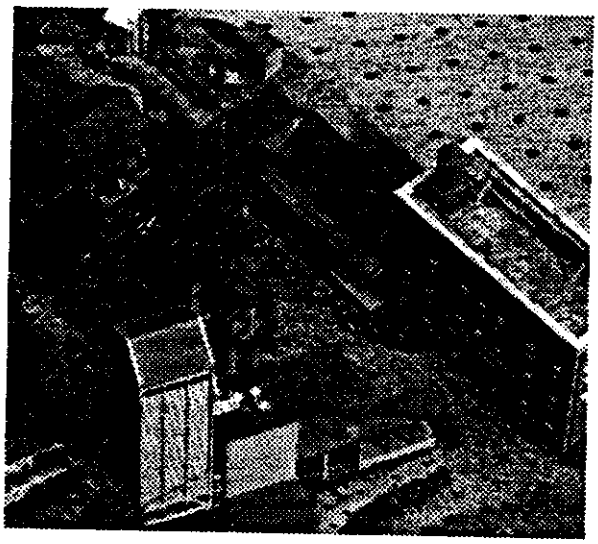
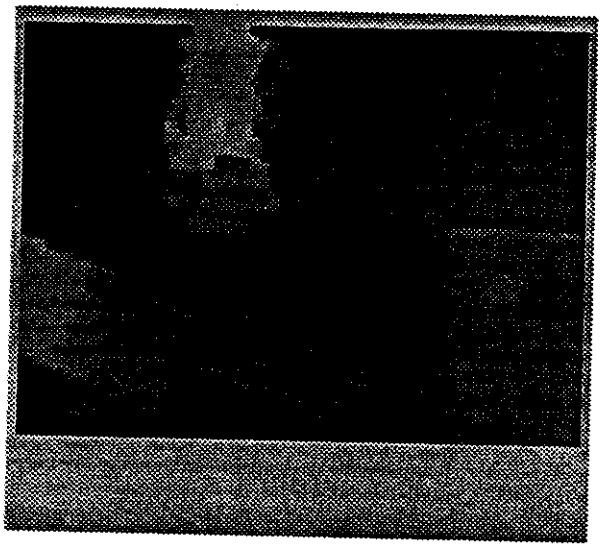
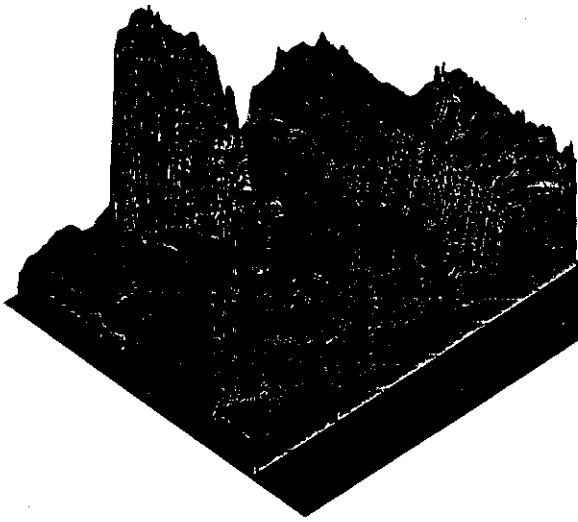


Figure 15: Oblique view

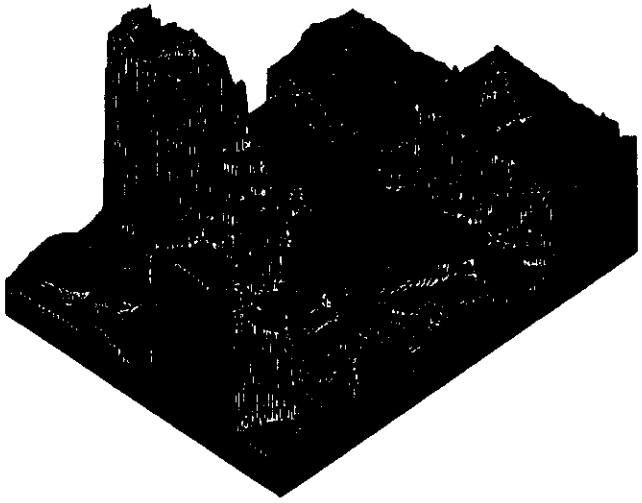


(a)

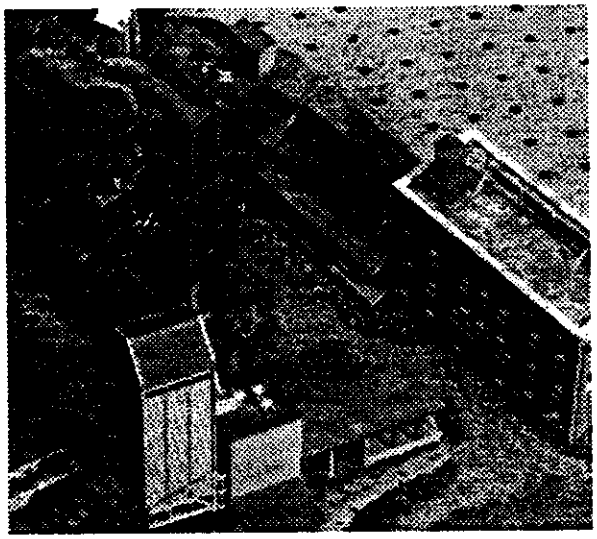


(b)

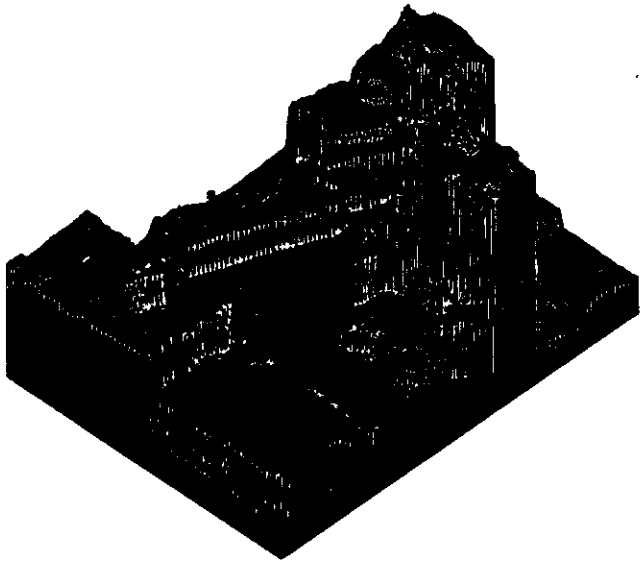
Figure 16: Multiple baselines: (a) Depth map; (b) Isometric plot



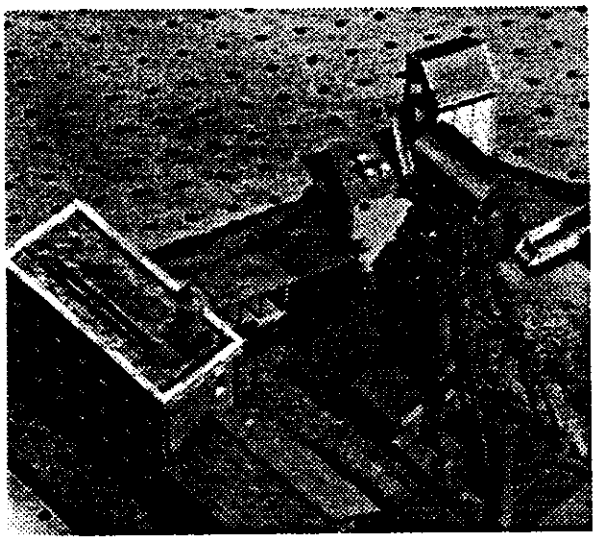
(a)



(b)



(c)



(d)

Figure 17: Isometric plots of the depth map computed by an iterative stereo algorithm with a locally adaptive window, in which the depth map of figure 16 from the multiple-baseline stereo was used as the initial map: (a) and (b) Isometric plot and corresponding view from the lower left corner; (c) and (d) Isometric plot and corresponding view from the upper right corner.

function, the SSSD-in-inverse-depth, exhibits an unambiguous and sharper minimum at the correct matching position. As a result there is no need for search or sequential estimation procedures.

The algorithm is easily amenable to parallel hardware implementation. Figure 4 shows a possible scheme with multiple cameras, SSD calculators, adders and time delays.

The key idea of the method is to relate SSD values to the inverse depth rather than the disparity. As an afterthought, this idea is natural. Whereas disparity is a function of the baseline, there is only one true (inverse) depth for each pixel position for all of the stereo pairs. Therefore there must be a single minimum for the SSD values when they are summed and plotted with respect to the inverse depth.

We have shown the advantage of the proposed method in removing ambiguity and improving precision 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.

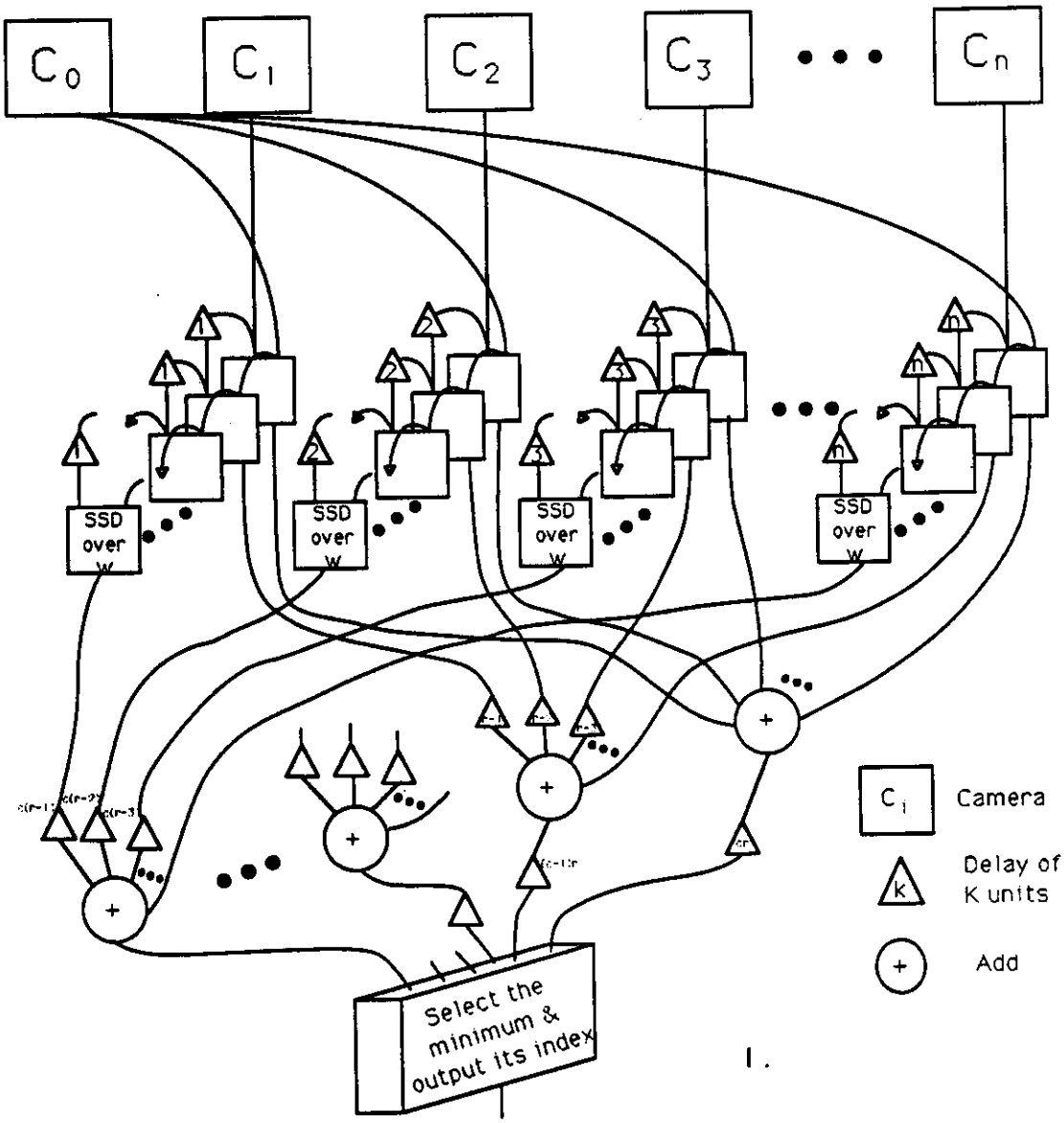


Figure 18: A scheme for parallel hardware implementation of the multiple baseline stereo algorithm by SSSD-in-inverse-depth. We place $n + 1$ multiple cameras along an axis. All cameras are synchronized. Pairs of camera output signals (the 0-th and the i-th) are fed into $nd_{(n)Max}$ SSD calculators with appropriate time delays, where $d_{(n)Max}$ is the maximum disparity for the stereo pair with the longest baseline. Each SSD calculator computes the SSD values between a particular image pair for a particular (quantized) inverse depth. The outputs of the SSD calculators for the same inverse depth are connected to an adder to produce the sum of SSD's (SSSD) for that inverse depth. These SSSD values are compared, and the inverse depth that shows the minimum is selected as the estimate.

A SSSD-in-inverse-depth for Ambiguous Pattern

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

$$f(x+j) = f(\xi+j), \qquad \text{if and only if } \xi = x \text{ or } \xi = x+a. \tag{32}$$

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

$$\begin{aligned} E[e_{\zeta(12)}(x,\zeta)] &= \sum_{j\in W}(f(x+j) - f(x+B_1F(\zeta-\zeta_r)+j))^2 \\ &+ \sum_{j\in W}(f(x+j) - f(x+B_2F(\zeta-\zeta_r)+j))^2 + 4N_w\sigma_n^2 \\ &> 4N_w\sigma_n^2 = E[e_{\zeta(12)}(x,\zeta_r)]. \end{aligned} \tag{33}$$

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

$$\sum_{j\in W}(f(x+j) - f(x+B_1F(\zeta_f-\zeta_r)+j))^2 + \sum_{j\in W}(f(x+j) - f(x+B_2F(\zeta_f-\zeta_r)+j))^2 = 0. \tag{34}$$

Then, it must be the case that

$$\begin{aligned} f(x+j) &= f(x+a_1+j) \\ \text{and}\quad f(x+j) &= f(x+a_2+j), \end{aligned} \tag{35}$$

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 $\zeta_r \neq \zeta_f$,

$$a_1 \neq a_2. \tag{36}$$

So, we have

$$f(x+j) = f(\xi+j), \qquad \text{for } \xi = x,\ x+a_1, \text{ or } x+a_2. \tag{37}$$

Since this contradicts assumption (32), equation (34) does not hold. Its left hand side must be positive. Hence (33) holds.

B Period of $E[e_{\zeta(i)}(x,\zeta)]$

Proposition: If $f(x)$ is a periodic function with a period T, then $E[e_{\zeta(i)}(x,\zeta)]$ is a periodic function with the period $\frac{T}{B_iF}$.

Proof: Let $f(x)$ be a periodic function with a period T. That is, for $\forall x$,

$$f(x) = f(x+kT), \qquad \text{for } k = 0, \pm 1, \pm 2, \ldots, \tag{38}$$

where $T > 0$ is the period, and there is no smaller value of T with this property. Using equations (13) and (38),

$$\begin{aligned} E[e_{\zeta(i)}\left(x, \zeta + k\frac{T}{B_iF}\right)] &= \sum_{j\in W}\left(f(x+j) - f\left(x + B_iF\left(\zeta + k\frac{T}{B_iF} - \zeta_r\right) + j\right)\right)^2 + 2N_w\sigma_n^2 \\ &= \sum_{j\in W}(f(x+j) - f(x + B_iF(\zeta - \zeta_r) + j + kT))^2 + 2N_w\sigma_n^2 \\ &= \sum_{j\in W}(f(x+j) - f(x + B_iF(\zeta - \zeta_r) + j))^2 + 2N_w\sigma_n^2 \\ &= E[e_{\zeta(i)}(x,\zeta)]. \end{aligned} \tag{39}$$

That is, $E[e_{\zeta(i)}(x,\zeta)]$ is a periodic function of ζ. Next, we show that $\frac{T}{B_iF}$ is its period. For that, tentatively suppose T' is the period such that

$$T' < \frac{T}{B_iF}. \tag{40}$$

Then,

$$E[e_{\zeta(i)}(x, \zeta + kT')] = E[e_{\zeta(i)}(x,\zeta)], \qquad \text{for } \forall\zeta. \tag{41}$$

Since this must be true for all ζ, let ζ be ζ_r. From equation (13).

$$\sum_{j\in W}(f(x+j) - f(x + kB_iT' + j))^2 = 0 \tag{42}$$

Therefore,

$$f(x) = f(x + kB_iFT'). \tag{43}$$

This means that $f(x)$ has a period B_iFT' that is shorter than T because of (40). This is a contradiction and thus T' cannot be the period of $E[e_{\zeta(i)}(x,\zeta)]$. Therefore, $E[e_{\zeta(i)}(x,\zeta)]$ is a periodic function of ζ with the period $\frac{T}{B_iF}$.

C Multiple-Baseline Stereo Algorithm

We present a complete description of the stereo algorithm using multiple-baseline stereo pairs. The task is, given n stereo pairs, find the ζ that minimizes the SSSD-in-inverse-depth function,

$$SSSD(x,\zeta) = \sum_{i=1}^{n} \sum_{j \in W} (f_0(x+j) - f_i(x + B_i F \zeta + j))^2. \tag{44}$$

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

Minimum of SSSD at Pixel Resolution

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

$$R_i = \frac{B_i}{B_n}. \tag{45}$$

Then,

$$B_i F \zeta = R_i B_n F \zeta = R_i d_{(n)}, \tag{46}$$

where $d_{(n)}$ is the disparity for the stereo pair with baseline B_n. Substituting this into equation (44),

$$SSSD(x, d_{(n)}) = \sum_{i=1}^{n} \sum_{j \in W} (f_0(x+j) - f_i(x + R_i d_{(n)} + j))^2. \tag{47}$$

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 baseline 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 by an iterative algorithm presented in [MO89][OK90]. 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-depth, 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:

$$\Delta\hat{d}_{(n)} = \frac{\sum_{j\in W}(f_0(x+j) - f_n(x+d_{0(n)}+j))f_n'(x+d_{0(n)}+j)}{\sum_{j\in W}(f_n'(x+d_{0(n)}+j))^2} \tag{48}$$

$$\sigma^2_{\Delta d_{(n)}} = \frac{2\sigma_n^2}{\sum_{j\in W}(f_n'(x+d_{0(n)}+j))^2}, \tag{49}$$

The value $\Delta\hat{d}_{(n)}$ 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

$$d_{0(n)} \leftarrow d_{0(n)} + \Delta\hat{d}_{(n)} \tag{50}$$

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

References

[BBM87] R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision*, 1(1):7–55, 1987.

[Gri85] W. E. L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(1):17–34, January 1985.

[Hee89] Joachim Heel. Dynamic motion vision. In *Proceedings of the DARPA Image Understanding Workshop*, pages 702–713, Palo Alto, Ca, May 23-26 1989.

[KO90] Takeo Kanade and Masatoshi Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. Technical Report CMU-CS-90-120, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, 1990.

[MO89] Larry Matthies and Masatoshi Okutomi. A bayesian foundation for active stereo vision. In *SPIE, Sensor Fusion II: Human and Machine Strategies*, November 1989.

[MP79] D. Marr and T. Poggio. A theory of human stereo vision. In *Proc. Roy. Soc. London*, pages 301–328, 1979.

[MSK89] 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.

[OK90] Masatoshi Okutomi and Takeo Kanade. A locally adaptive window for signal matching. In *Proc. Third Int' l Conference on Computer Vision*, December 1990.

[Yam88] Masanobu Yamamoto. The image sequence analysis of three-dimensional dynamic scenes. Technical Report 893, Electrotechnical Laboratory - Agency of Industrial Science and Technology, Tsukuba, Ibaraki, Japan, May 1988.