Proceedings of the 1991 IEEE
International Conference on Robotics and Automation
Sacramento, California - April 1991

A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment *

Takeo Kanade
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA

Masatoshi Okutomi†
Information Systems Research Center
Canon Inc.
890-12 Kashimada, Saiwai, Kawasaki, 211, Japan

Abstract

A central problem in stereo matching by computing correlation or sum of squared differences (SSD) lies in selecting an appropriate window size. The window size must be large enough to include enough intensity variation for matching but small enough to avoid the effects of projective distortion. If the window is too small and does not cover enough intensity variation, it gives a poor disparity estimate, because the signal (intensity variation) to noise ratio is low. If, on the other hand, the window is too large and covers a region in which the depth of scene points (ie., disparity) varies, then the position of maximum correlation or minimum SSD may not represent correct matching due to different projective distortions in the left and right images. For this reason, a window size must be selected adaptively depending on local variations of intensity and disparity.

The stereo algorithm we present selects a window adaptively by evaluating the local variation of the intensity and the disparity. We employ a statistical model that represents uncertainty of disparity of points over the window: the uncertainty is assumed to increase with the distance of the point from the center point. This modeling enables us to assess how disparity variation within a window affects the estimation of disparity. As a result, we can compute the uncertainty of the disparity estimate which takes into account both intensity and disparity variances. So, the algorithm can search for a window that produces the estimate of disparity with the least uncertainty for each pixel of an image. The method controls not only the size but also the shape (rectangle) of the window. The algorithm has been tested on both synthetic and real images, and the quality of the disparity maps obtained demonstrates the effectiveness of the algorithm.

1 Introduction

Stereo matching by computing correlation or sum of squared differences (SSD) is a basic technique for obtaining a dense depth map from images [6][2][3][12][10][7]. As Barnard and Fischler [1] point out, "a problem with correlation (or SSD) matching is that the patch (window) size must be large enough to include enough intensity variation for matching but small enough to avoid the effects of projective distortion." If the window is too small and does not cover enough intensity variation, it gives a poor disparity estimate, because the signal (intensity variation) to noise ratio is low. If, on the other hand, the window is too large and covers a region in which the depth of scene points (ie., disparity) varies, then the position of maximum correlation or minimum SSD may not represent correct matching due to different projective distortions in the left and right images. For this reason, a window size must be selected adaptively depending on local variations of intensity and disparity.

However, most correlation- or SSD-based stereo methods in the past have used a window of a fixed size that is chosen empirically for each application. There has been little research for adaptive window selection. As a relevant technique, Panton [10] warped the image to account for predicted terrain relief, but failed to consider contribution due to intensity variation. Levine et. al [4] changed the window size locally depending on the intensity pattern, but uncertainty in matching due to the variation of unknown disparities was unaccounted for.

The difficulty of a locally adaptive window lies, in fact, in a difficulty in evaluating and using disparity variances. While the intensity variation is directly obtained from the image, evaluation of the disparity variation is not easy, since the disparity *is* what we intend to calculate as an end product of stereo. To resolve the dilemma, an appropriate model of disparity variation is required which enables us to assess how disparity variation within a window affects the estimation of disparity.

The stereo algorithm we propose in this paper selects a window adaptively by evaluating the local variation of the intensity and the disparity. We employ a statistical model that represents uncertainty of disparity of points over the window: the uncertainty is assumed to increase with the distance of the point from the center point. This modeling enables us to compute both a disparity estimate *and* the uncertainty of the estimate. So, the algorithm can search for a window that produces the estimate of disparity with the least uncertainty for each pixel of an image. The method controls not only the size but also the shape (rectangle) of the window.

In this paper, we first develop a model of stereo matching in section 2. Section 3 shows how to estimate the most likely disparity and the uncertainty of the estimate based on the modeling in section 2. These two sections provide theoretical grounds of our proposed algorithm. In section 4, we presents a complete stereo algorithm which selects appropriate window size and shape adaptively for each pixel. Section 5 provides experimental results with real stereo images. The quality of the disparity maps obtained demonstrates the effectiveness of the algorithm.

*This research was supported 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 author and should not be interpreted as representing the official policies, either expressed or implied, of DARPA or the U.S. government.

†This research was performed while the second author was with Carnegie Mellon University.

2 Modeling Stereo Matching

We will first develop a statistical model of the difference of intensities of two images within a window. The analysis is based on the uncertainty model presented in [8]. Let the stereo intensity images be $f_1(x,y)$ and $f_2(x,y)$. Assume that the baseline is parallel to the x-axis, and $f_1(x,y)$ and $f_2(x,y)$ come from the same underlying intensity function with a disparity function $d_r(x,y)$ and additive noise. Then f_1 and f_2 are related by

$$f_1(x,y) = f_2(x + d_r(x,y), y) + n(x,y), \tag{1}$$

where $n(x,y)$ is Gaussian white noise

$$n(x,y) \sim N(0, 2\sigma_n^2). \tag{2}$$

The value σ_n^2 is the power of noise per image.*

To simplify the notation, suppose that we want to compute the disparity at $(x,y) = (0,0)$, i.e., the value $d_r(0,0)$. Also, suppose a window $W = \{(\xi,\eta)\}$ is placed at the correct corresponding positions in both images, that is, at $(0,0)$ in image $f_1(x,y)$ and at $(d_r(0,0),0)$ in image $f_2(x,y)$. Figure 1 illustrates the situation. Then, the value of f_1 at (ξ,η) in the window is $f_1(\xi,\eta)$, and that of f_2 is $f_2(\xi + d_r(0,0),\eta)$. These values would be the same, except the noise component, if the disparity $d_r(\xi,\eta)$ were constant and equal to $d_r(0,0)$, but in general they are not. By expanding $f_2(\xi + d_r(\xi,\eta),\eta)$ at $\xi + d_r(0,0)$ and using equation (1), we see that the difference of intensities between f_1 and f_2 at (ξ,η) in the window can be approximated as

$$\begin{aligned} & f_1(\xi,\eta) - f_2(\xi + d_r(0,0),\eta) \\ \approx\ & (d_r(\xi,\eta) - d_r(0,0))\frac{\partial}{\partial\xi} f_2(\xi + d_r(0,0),\eta) + n(\xi,\eta). \end{aligned} \tag{3}$$

At this point, let us introduce the following statistical model for the disparity $d_r(\xi,\eta)$ within a window:

$$d_r(\xi,\eta) - d_r(0,0) \sim N\left(0, \alpha_d\sqrt{\xi^2+\eta^2}\right), \tag{4}$$

where α_d is a constant that represents the amount of fluctuation of the disparity. That is, this model assumes that the difference of disparity at a point (ξ,η) in the window from that of the center point $(0,0)$ has a zero-mean Gaussian distribution with variance proportional to the distance between these points. In other words, the expected value of the disparity at (ξ,η) is the same as the center point, but it is expected to fluctuate more as the point is farther from the center.† Or, in terms of the scene, the surface covered by the window is expected to be locally flat and parallel to the baseline, but it is less certain as the window becomes larger. We also assume that the image intensity derivatives $\frac{\partial}{\partial\xi} f_2(\xi,\eta)$ within a window follow a zero-mean Gaussian white distribution,‡ and that intensity derivatives $\frac{\partial}{\partial\xi} f_2(\xi,\eta)$ and disparities $d_r(\xi,\eta)$ are mutually independent.

*We use $2\sigma_n^2$ in equation (2) as variance of $n(x,y)$ to indicate that it includes noise added to both f_1 and f_2.

†The statistical model of (4) can be shown equivalent to assuming that $d_r(\xi,\eta)$ is generated by Brownian motion (refer to [5][11]). More generally, we can assume $d_r(\xi,\eta)$ to be a fractal. This corresponds to choosing a different degree of $\xi^2+\eta^2$ in the variance in (4). The Brownian motion is the simplest case in which the degree is $\frac{1}{2}$. However, our preliminary experiments have shown no noticeable advantage of using a general fractal assumption.

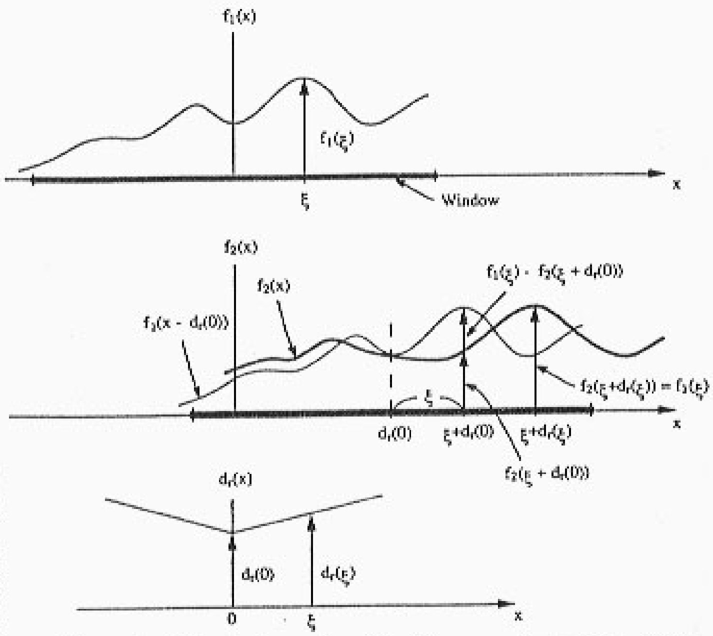


Figure 1: Illustration of $n_s(\xi,\eta)$ in one dimension. The graph at the top shows $f_1(x)$; the middle one, $f_2(x)$ (the thicker curve) with $f_1(x)$ shifted by $d_r(0)$ (the thinner curve); the bottom one, $d_r(x)$. The region indicated by the very thick lines on the axes indicate the region covered by the window.

These assumptions allow us to model a statistical distribution of the intensity difference (3). Let us denote the right hand side of equation (3) by $n_s(\xi,\eta)$. First, we compute the mean and variance of $n_s(\xi,\eta)$:

$$\begin{aligned} & E\left[n_s(\xi,\eta)\right] \\ =\ & E[d_r(\xi,\eta) - d_r(0,0)]E\left[\frac{\partial}{\partial\xi} f_2(\xi + d_r(0,0),\eta)\right] \\ & + E[n(\xi,\eta)] = 0 \end{aligned} \tag{5}$$

$$\begin{aligned} & E\left[(n_s(\xi,\eta))^2\right] \\ =\ & E\left[\left((d_r(\xi,\eta) - d_r(0,0))\frac{\partial}{\partial\xi} f_2(\xi + d_r(0,0),\eta)\right)^2\right] \\ & + E\left[2(d_r(\xi,\eta) - d_r(0,0))\cdot \right. \\ & \left. \cdot\left(\frac{\partial}{\partial\xi} f_2(\xi + d_r(0,0),\eta)\right) n(\xi,\eta)\right] + E\left[(n(\xi,\eta))^2\right] \\ =\ & E\left[(d_r(\xi,\eta) - d_r(0,0))^2\right]\cdot \end{aligned}$$

‡This is also equivalent to assuming the pattern $f_2(\xi,\eta)$ to be result of Brownian motion: i.e., locally it has a constant brightness, but has more fluctuation as the window becomes bigger.

$$\cdot E\left[\left(\frac{\partial}{\partial\xi}f_2(\xi+d_r(0,0),\eta)\right)^2\right]+E\left[(n(\xi,\eta))^2\right]$$

$$= \quad 2\sigma_n^2+\alpha_f\alpha_d\sqrt{\xi^2+\eta^2}, \tag{6}$$

where

$$\alpha_f = E\left[\left(\frac{\partial}{\partial\xi}f_2(\xi+d_r(0,0),\eta)\right)^2\right]. \tag{7}$$

We can show that $n_s(\xi,\eta)$ is white noise and its distribution can be approximated by a Gaussian distribution with the above mean and variance. That is,

$$\begin{aligned} n_s(\xi,\eta) &\approx f_1(\xi,\eta)-f_2(\xi+d_r(0,0),\eta) \\ &\sim N\left(0,2\sigma_n^2+\alpha_f\alpha_d\sqrt{\xi^2+\eta^2}\right). \end{aligned} \tag{8}$$

The intuitive interpretation of (8) is as follows. Referring to figure 1, $n_s(\xi,\eta)$ is the difference between f_1 and f_2 at (ξ,η) within a window when the window is placed at the corresponding positions for obtaining the disparity at $(0,0)$. If there is no additive noise $n(x,y)$ in the image (i.e., $\sigma_n=0$) and the disparity is constant within the window (i.e., $\alpha_d=0$), then the two images match exactly, and $n_s(\xi,\eta)$ must be null. Otherwise, however, the difference has a value which shows a combined noise characteristic which comes from both intensity and disparity variations. As derived in (8), it can be modeled by zero-mean Gaussian noise whose variance is a summation of a constant term and a term proportional to $\sqrt{\xi^2+\eta^2}$. The constant term is from the noise added to the image intensities. The second term is from *uncertain local support*. That is, while the points surrounding the center point in the window are used to support the matching for the center point, it should be noted that these points may actually increase the error in computing the disparity of the center point. This is because, in general, the disparity of the surrounding points deviates from that of the center point. This uncertainty is represented as if the intensity signals have additional noise whose power is proportional to the distance from the center point in the window. If the disparity is constant over the window (i.e. $\alpha_d=0$), the additional noise is zero. If the disparity changes more in the window (i.e., the larger α_d is), its effect becomes larger and the information contributed by the surrounding points becomes more uncertain.

3 Estimating Disparity and Its Uncertainty

Now, we will show how the disparity and its uncertainty can be estimated based on the modeling presented in the previous section. Let $d_0(x,y)$ be an initial estimate of the disparity $d_r(x,y)$. By using the Taylor expansion, equation (8) becomes

$$\begin{aligned} & f_1(\xi,\eta)-f_2(\xi+d_0(0,0),\eta)-\Delta d\frac{\partial}{\partial\xi}f_2(\xi+d_0(0,0),\eta) \\ =\ & n_s(\xi,\eta), \end{aligned} \tag{9}$$

where Δd is an incremental correction of the estimate to be made, such that $\Delta d=d_r(0,0)-d_0(0,0)$. Dividing both sides of this equation by $\sqrt{2\sigma_n^2+\alpha_f\alpha_d\sqrt{\xi^2+\eta^2}}$ yields

$$\begin{aligned} & \frac{f_1(\xi,\eta)-f_2(\xi+d_0(0,0),\eta)-\Delta d\frac{\partial}{\partial\xi}f_2(\xi+d_0(0,0),\eta)}{\sqrt{2\sigma_n^2+\alpha_f\alpha_d\sqrt{\xi^2+\eta^2}}} \\ =\ & n_n(\xi,\eta), \end{aligned} \tag{10}$$

where $n_n(\xi,\eta)=\frac{n_s(\xi,\eta)}{\sqrt{2\sigma_n^2+\alpha_f\alpha_d\sqrt{\xi^2+\eta^2}}}$ is Gaussian white noise such that

$$n_n(\xi,\eta)\sim N(0,1). \tag{11}$$

By letting

$$\phi_1(\xi,\eta) = \frac{f_1(\xi,\eta)-f_2(\xi+d_0(0,0),\eta)}{\sqrt{2\sigma_n^2+\alpha_f\alpha_d\sqrt{\xi^2+\eta^2}}} \tag{12}$$

$$\phi_2(\xi,\eta) = \frac{\frac{\partial}{\partial\xi}f_2(\xi+d_0(0,0),\eta)}{\sqrt{2\sigma_n^2+\alpha_f\alpha_d\sqrt{\xi^2+\eta^2}}}, \tag{13}$$

we have

$$\phi_1(\xi,\eta)-\Delta d\phi_2(\xi,\eta) = n_n(\xi,\eta). \tag{14}$$

Now, by sampling ϕ_1 and ϕ_2 at (ξ_i,η_j) in the window W we can define ξ_{ij} as

$$\xi_{ij} = \phi_1(\xi_i,\eta_j)-\Delta d\phi_2(\xi_i,\eta_j) = n_n(\xi_i,\eta_j). \tag{15}$$

From equation (11), the conditional probability density function of ξ_{ij} given Δd is

$$p(\xi_{ij}|\Delta d) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\phi_1(\xi_i,\eta_j)-\Delta d\phi_2(\xi_i,\eta_j))^2}{2}\right). \tag{16}$$

Since $n_n(\xi,\eta)$ is white noise, ξ_{ij} are mutually independent. So we get

$$p(\xi_{ij}(i,j\in W)|\Delta d) = \prod_{i,j\in W}p(\xi_{ij}|\Delta d), \tag{17}$$

where $p(\xi_{ij}(i,j\in W)|\Delta d)$ is the conditional joint probability for the points in the window, and $\prod_{i,j\in W}$ denotes the product over the window. From the continuous version of Bayes' theorem,

$$p(\Delta d|\xi_{ij}(i,j\in W)) = \frac{p(\xi_{ij}(i,j\in W)|\Delta d)p(\Delta d)}{\int_{-\infty}^{\infty}p(\xi_{ij}(i,j\in W)|\Delta d)p(\Delta d)d(\Delta d)}. \tag{18}$$

Assuming no prior information of Δd (i.e., $p(\Delta d)=$ *constant*), substitution of (16) into (18) yields

$$p(\Delta d|\xi_{ij}(i,j\in W)) = \frac{1}{\sqrt{2\pi}\sigma_{\Delta d}}\exp\left(-\frac{(\Delta d-\hat{\Delta d})^2}{2\sigma_{\Delta d}^2}\right), \tag{19}$$

where

$$\hat{\Delta d} = \frac{\sum_{i,j\in W}(\phi_1(\xi_i,\eta_j)\phi_2(\xi_i,\eta_j))}{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j))^2} \tag{20}$$

$$\sigma_{\Delta d}^2 = \frac{1}{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j))^2}, \tag{21}$$

where $\sum_{i,j\in W}$ denotes the summation over the window. Equation (19) says that the conditional probability density function of Δd given the observed stereo intensities over the window becomes a Gaussian probability density function. The mean value and the variance of the Gaussian probability are $\hat{\Delta d}$ and $\sigma_{\Delta d}^2$, computed with equations (20) and (21). That is, $\hat{\Delta d}$ and $\sigma_{\Delta d}^2$ provide the maximum likelihood estimate of the disparity (increment) and the uncertainty of the estimation for the given window W, respectively.

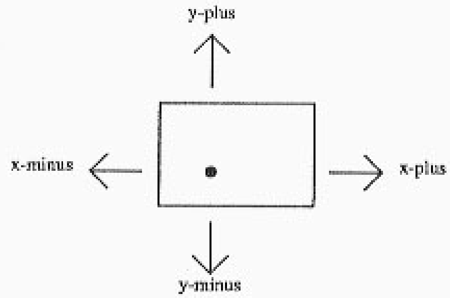


Figure 2: Window expansion

α_d and α_f are parameters that represent the disparity fluctuation and the intensity fluctuation, respectively. We estimate them locally within the window from equations (4) and (7),

$$\hat{\alpha_d} = \frac{1}{N_w} \sum_{i,j \in W} \frac{(d_0(\xi_i, \eta_i) - d_0(0,0))^2}{\sqrt{\xi_i^2 + \eta_i^2}} \quad (22)$$

$$\hat{\alpha_f} = \frac{1}{N_w} \sum_{i,j \in W} \left(\frac{\partial}{\partial \xi} f_2(\xi_i + d_0(0,0), \eta_i) \right)^2, \quad (23)$$

where N_w is the number of the samples within the window. These parameters change as the shape and size of a window changes.

4 Iterative Stereo Algorithm with an Adaptive Window

In the previous sections we have developed a theory for computing the estimates of the disparity increment and its uncertainty, which take into account the fact that not only the intensity but also the disparity varies within a window. We now describe the complete stereo algorithm based on the theory:

1. Start with an initial disparity estimate $d_0(x, y)$. This initial estimate can be obtained by any existing stereo algorithm.
2. For each point (x, y), choose a window that provides the estimate of disparity increment having the lowest uncertainty. For the chosen window, calculate the disparity increment by (20) and update the disparity estimate by $d_{i+1}(x, y) = d_i(x, y) + \Delta d(x, y)$.

   Here we need a strategy to select a window that results in the disparity estimate having the lowest uncertainty. In the discussions so far the shape of the window can be arbitrary. In practice we limit ourselves to a rectangular window, as illustrated in figure 2, whose width and height can be independently controlled in all four directions. Our strategy is as follows:

   (a) Place a small 3×3 window centered at the pixel, and compute the uncertainty by using (22), (23), and (21).

   (b) Expand the window by one pixel in one direction, e.g., to the right $x+$, for trial, and compute the uncertainty for the expanded window. If the expansion increases the uncertainty, the direction is prohibited from further expansions. Repeat the same process for each of the four directions $x+, x-, y+$, and $y-$ (excluding the already prohibited ones).

   (c) Compare the uncertainties for all the directions tried and choose the direction which produces the minimum uncertainty.

   (d) Expand the window by one pixel in the chosen direction.

   (e) Iterate steps (b) to (d) until all directions become prohibited from expansion or until the window size reaches to a limit that is previously set.

   Thus, our strategy is basically a sequential search for the best window by maximum descent starting with the smallest window
3. Iterate the above process until the disparity estimate $d_i(x, y)$ converges, or up to a certain maximum number of iterations.

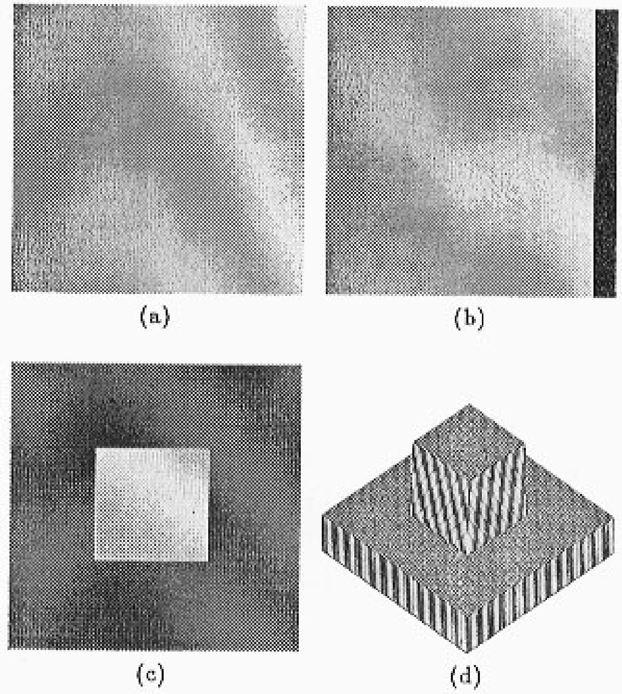


Figure 3: Synthesized stereo images, with a ramp intensity pattern with Gaussian noise: (a) Left image; (b) Right image; (c) Disparity pattern; (d) An isometric plot of the disparity pattern

Now, by using synthesized data we will examine how the window is adaptively set by the stereo algorithm for each position in an image, and demonstrate its advantage. Figures 3 (a) and (b) show the left and the right images of the test data. In generating the data set, a linear ramp in the direction of the baseline is used as the underlying intensity pattern. It is deformed according to the disparity pattern in figures 3 (c) and (d), and Gaussian noise is added independently to both images. We apply the iterative stereo algorithm to the resultant data.

First, we will examine the result of window selection at several representative positions shown in figure 4. The windows selected at those positions are drawn by dashed lines in figure 5 relative to the disparity edges drawn by solid

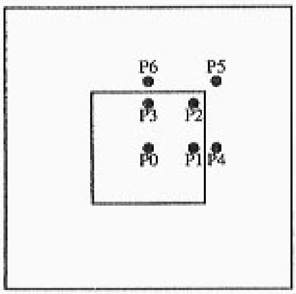


Figure 4: Positions for which size and shape of selected windows are examined.

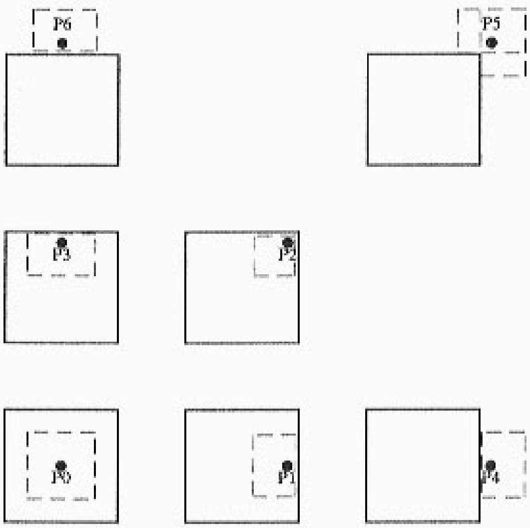


Figure 5: Selected windows for each position

lines.[§] For example, at $P0$ a window has been expanded to the limit for all directions, whereas at $P1$ expansion to the right has been stopped at the disparity edge. At $P5$, a window is elongated either vertically or horizontally, depending on the image noise, but consistently avoids the corner of the disparity jump.

Next, let us examine the computed disparities. For comparison, we also have computed disparities by running the same iterative algorithm but with a fixed window size; that is, in Step 2 of the stereo algorithm we use a window of predetermined size rather than the window selection strategy. We run with two window sizes, 3×3 and 7×7. Figures 6 (a) and (b) show the result produced by fixed window sizes, and (c) by the adaptive-window algorithm. We can clearly see the problem with using a predetermined fixed window size. A larger window is good for flat surfaces, but it blurs the disparity edges. In contrast, a smaller window gives sharper disparity edges at the expense of noisy surfaces. The computed disparity by the proposed algorithm shown in figure 6 (c) shows both smooth flat surfaces and sharp disparity edges. The improvements are further visible by plotting the absolute difference between the computed and true disparities as shown in figure 7, with a table that lists their mean error values. The adaptive-window algorithm has the smallest mean error, but more importantly we should observe that the algorithm has reduced two types of errors. A small fixed window results in large random error everywhere. A large fixed window removes the random error, but introduces systematic errors along the disparity edges. The adaptive-window based method generates small errors of both types. In fact, we have shown that at *each* point the expected value of the error by the adaptive-window method is always smaller than or equal to that produced when *any* fixed-size window is used [8].

[§]Actually these are the average of ten runs with different noise to obtain the general tendency, rather than accidental set up.

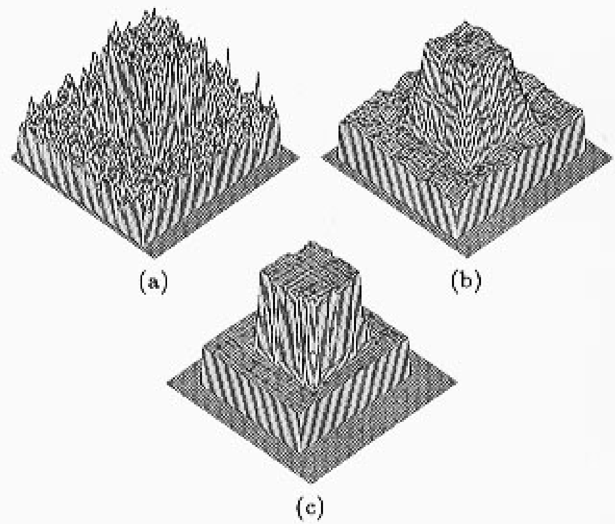
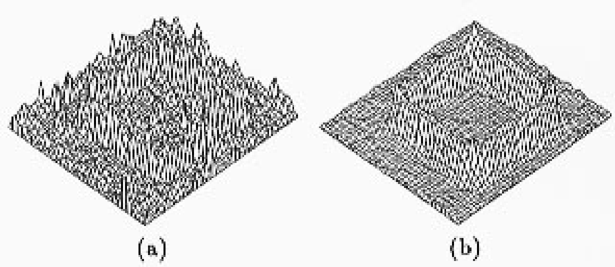
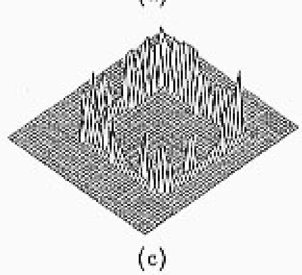


Figure 6: Isometric plots of the computed disparity by: (a) a 3×3 window; (b) a 7×7 window; (c) the adaptive window algorithm.





Window	Mean Error Value (pixel)
3×3	0.22
7×7	0.20
Adaptive Window	0.08

Figure 7: Difference between the true disparity and the computed disparity: (a) by a 3×3 window; (b) by a 7×7 window; (c) by the adaptive window.

5 Experimental Results

We have applied the adaptive-window based stereo matching algorithm presented in this paper to real stereo images.

Figure 8 shows images of a town model that were taken by moving the camera vertically. The disparity, therefore, is in the vertical direction.

For initial disparity estimates, we have used a technique of multiple-baseline stereo matching [9] which can remove matching ambiguities due to repetitive patterns, especially in the top portion of the image. Figure 9 (a) shows the final disparity map computed by the adaptive window algorithm. In addition, the uncertainty estimate computed by the algorithm is shown in figure 9 (b): increasing brightness corresponds to higher uncertainty. With this uncertainty estimate we can locate the regions whose computed disparity is not very reliable (very white regions in figure 9 (b)). In this example, they are either due to aliasing caused by the fine texture of roof tiles of a building (in the middle part of the image) or due to occlusion (the others). The disparity estimates of those uncertain parts can be discarded for later processing.

Figure 10 shows perspective views of the recovered scene by texture mapping the original intensity image on the constructed depth map and generating views from new positions which are outside of the original stereo views. They can give an idea of the quality of reconstruction. This stereo data set is the same one used in [6]. We can observe a noticeable improvement of the result over the previous result. Also it should be noted that this is extremely narrow baseline stereo: the baseline is only 1.2 cm long and the scene is about 1m away from the camera, thus the depth to the baseline ratio is approximately 80.

Figures 11 (a) and (b) show another set of real stereo images which are top views of a coal mine model. Figure 12 (a) shows the isometric plot of the computed disparity. For comparison, an actual picture of the model taken from roughly the same angles is given in figure 12 (b). The shapes of buildings, a $\wedge$-shaped roof, a water tank on the roof, and a flat ground have been recovered without blurring edges.

6 Conclusions

In this paper, we have presented an iterative stereo matching algorithm using an adaptive window. The algorithm selects a window adaptively for each pixel. The selected window is optimal in the sense that it produces the disparity estimate having the least uncertainty. By evaluating both the intensity and the disparity variations within a window, we can compute both the disparity estimate and its uncertainty which can then be used for selecting the optimal window.

The key idea for the algorithm is that it employs a statistical model that represents uncertainty of disparity of points over the window: the uncertainty is assumed to increase with the distance of the point from the center point. This model has enabled us to assess how disparity variation within a window affects the estimation of disparity.

An important feature of the algorithm is that it is completely local and does not include any global optimization. Also, the algorithm does not use any post-processing smoothing, but smooth surfaces are recovered as smooth while sharp disparity edges are retained.

The experimental results have demonstrated a clear advantage of this algorithm over algorithms with a fixed-size window both on synthetic and on real images.

References

[1] S.T. Barnard and M.A. Fischler. *Stereo Vision*, pages 1083–1090. Wiley, New York, 1987.

[2] Wolfgang Forstner and Alfred Pertl. *Photogrammetric Standard Methods and Digital Image Matching Techniques for High Precision Surface Measurements*, pages 57–72. Elsevier Science Publishers B.V., 1986.

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

[4] Martin D. Levine, Douglas A. O'Handley, and Gary M. Yagi. Computer determination of depth maps. *Computer Graphics and Image Processing*, 2(4):131–150, 1973.

[5] B. B. Mandelbrot and B. J. Van Ness. Fractional brownian motion, fractional noises and applications. *SIAM*, 10(4):422–438, 1968.

[6] 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.

[7] Kenichi Mori, Masatsugu Kidode, and Haruo Asada. An iterative prediction and correction method for automatic stereo comparison. *Computer Graphics and Image Processing*, 2:393–401, 1973.

[8] Masatoshi Okutomi and Takeo Kanade. A locally adaptive window for signal matching. Technical Report CMU-CS-90-178, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, 1990. Also appeared as A Locally Adaptive Window for Signal Matching. Proc. of Int'l Conf. on Computer Vision, December, 1990.

[9] Masatoshi Okutomi and Takeo Kanade. A multiple-baseline stereo. Technical Report CMU-CS-90-189, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, 1990.

[10] D. J. Panton. A flexible approach to digital stereo mapping. *Photogram. Eng. Remote Sensing*, 44(12):1499–1512, Dec 1978.

[11] Richard F. Voss. Fractals in nature. In *Course note on FRACTALS: Introduction, Basics, and Perspectives*, 1987.

[12] G.A. Wood. Realities of automatic correlation problem. *Photogrammetric Engineering and Remote Sensing*, 49:537–538, April 1983.

(a) (b)

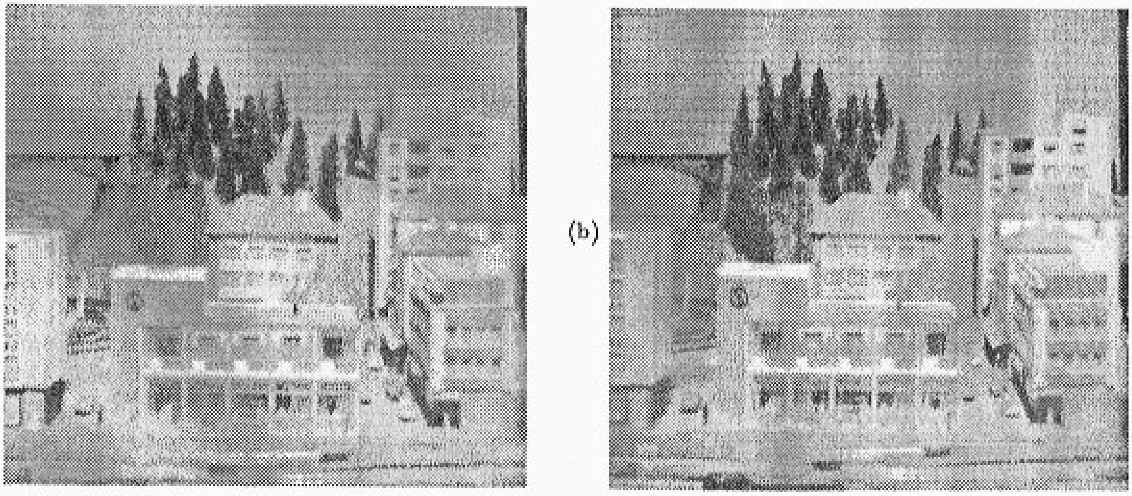


Figure 8: "Town" stereo data set: (a) Upper image of stereo;(b) Lower image of stereo.

(a) (b)

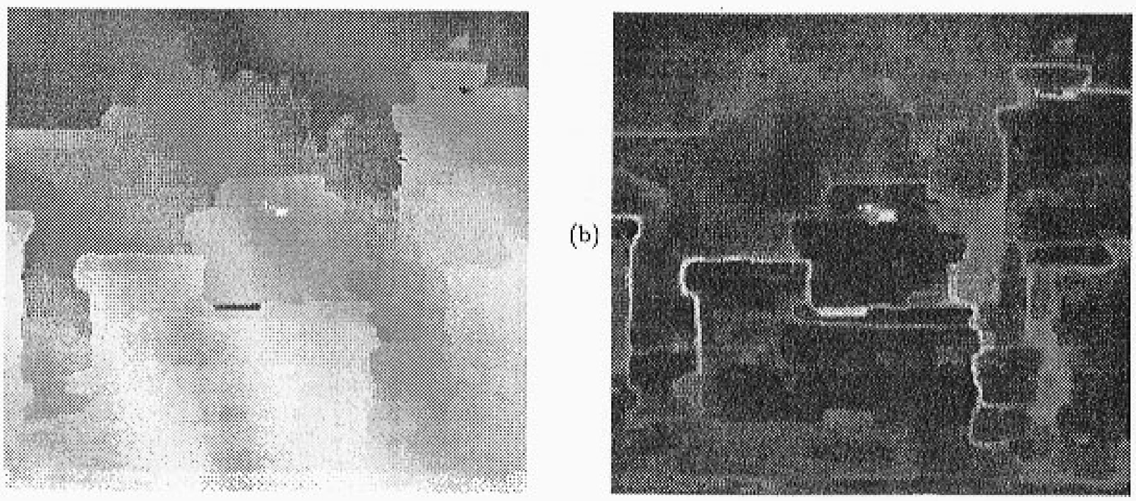


Figure 9: Computed disparity and uncertainty for the "town" stereo data: (a) Disparity map; (b) Uncertainty;

(a) (b)

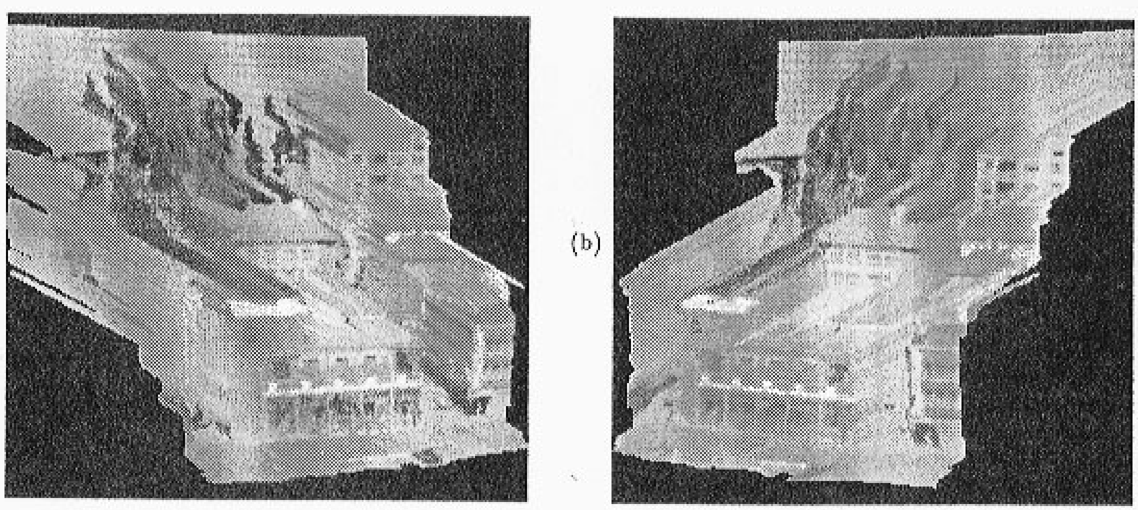
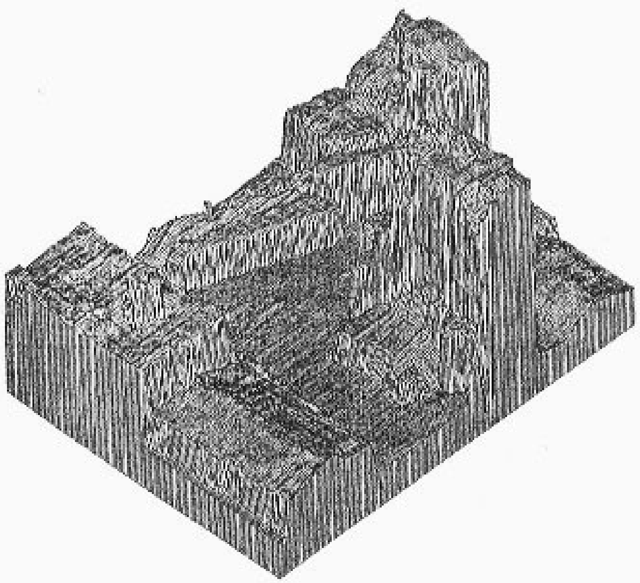


Figure 10: Perspective views of the recovered scene: (a) from an upper left position; (b) from an upper right position.

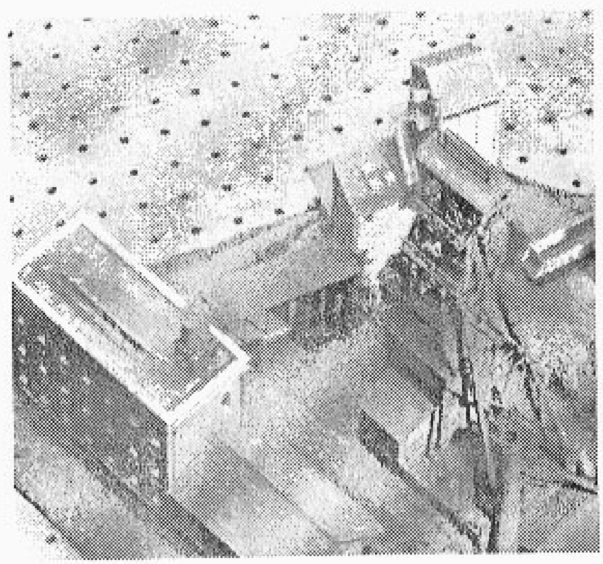
(a)

(b)

Figure 11: "Coal mine" stereo data set: (a) Lower image; (b) Upper image.



(a)



(b)

Figure 12: Isometric plots of the computed disparity map and their corresponding actual view: (a) (b) Isometric plot and corresponding view from the upper right corner.