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

Takeo Kanade, *Fellow, IEEE*, and Masatoshi Okutomi

***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 reliable 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 (i.e., 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.**

We present a method to select an appropriate window by evaluating the local variation of the intensity *and* the disparity. We employ a statistical model of the disparity distribution within the window. This modeling enables us to assess how disparity variation, as well as intensity variation, within a window affects the uncertainty of disparity estimate at the center point of the window. As a result, we can devise a method which searches 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.

We have embedded this adaptive-window method in an iterative stereo matching algorithm: starting with an initial estimate of the disparity map, the algorithm iteratively updates the disparity estimate for each point by choosing the size and shape of a window till it converges. The stereo matching algorithm has been tested on both synthetic and real images, and the quality of the disparity maps obtained demonstrates the effectiveness of the adaptive window method.

***Index Terms*—Stereo vision, adaptive window, statistical model, uncertainty, 3-D vision.**

I. INTRODUCTION

STEREO matching by computing correlation or sum of squared differences (SSD) is a basic technique for obtaining a dense depth map from images [13], [6], [9], [23], [18], [14]. 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 (i.e., 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 [18] warped the image to account for predicted terrain relief, but failed to consider the contribution due to intensity variation. In their coarse-to-fine stereo technique, Hoff and Ahuja [10] discuss the relationship between window shape and disparity, and argue how integrating the processes of matching, contour detection and surface interpolation can help reduce the problem. Levine *et al.* [11] 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 measurable 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.

We propose to employ a statistical model of the disparity distribution within the window: the difference of disparity at a point 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. This modeling enables us to compute the uncertainty of the disparity estimate by taking into account both intensity and disparity variances. As a result, we can devise a method which searches for a window that produces the estimate of disparity with the least uncertainty

Manuscript received July 20, 1990; revised January 24, 1994. 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. This work was performed while M. Okutomi was with Carnegie Mellon University. Recommended for acceptance by Associate Editor N. Ahuja.

T. Kanade is with the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213 USA; e-mail: tk@cs.cmu.edu.

M. Okutomi is with the Department of Control and Systems Engineering, Tokyo Institute of Technology, 2–12–1 O-okayama, Meguro-ku, Tokyo, 152 Japan; e-mail: mxo@ctrl.titech.ac.jp.

IEEE Log Number 9403147.

for each pixel of an image: the method controls not only the size but also the shape (rectangle) of the window.

We have embedded this adaptive-window method in an iterative stereo matching algorithm: starting with an initial estimate of the disparity map, the algorithm iteratively updates the disparity estimate for each point by choosing the size and shape of a window till it converges. The stereo matching algorithm has been tested on both synthetic and real images, and the quality of the disparity maps obtained demonstrates the effectiveness of the adaptive window method.

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

II. Modeling Stereo Matching

We will first develop a statistical model of the distribution of the difference of intensities of two images within a window. The analysis is based on the disparity distribution model presented in [15]. Then we will compare our model of the disparity distribution with local support models of other stereo methods.

A. Distributions of Intensity Differences and Disparities in a Window

Let the stereo intensity images (or results of some preprocessing) be $f_1(x,y)$ and $f_2(x,y)$. Without loss of generality, we can assume that the baseline is parallel to the x-axis. Further let us assume $f_1(x,y)$ 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 the noise per image.[1]

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 for 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 $\xi - d_r(0,0)$ up to the linear term and using (1), we see that the difference of intensities between f_1 and f_2 at (ξ,η) in the window can be approximated as

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

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 in 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.[2] Or, in terms of the scene, the small surface corresponding to the window in the image is statistically expected to be locally flat and parallel to the baseline, but the expectation becomes less certain as the window becomes larger. More discussions on the implication of this assumption are presented in Section II-B.

In order to facilitate the mathematical derivations to follow, we make an additional assumption that the image intensity $f_2(\xi,\eta)$ within the window is also generated by another Brownian process which is independent of the one that has generated $d_r(\xi,\eta)$. This means that the image intensity at a point within a window is expected to be the same as the center point, but that expectation is less certain the farther the point is from the center. In terms of the distribution of image intensity derivatives $\frac{\partial}{\partial\xi} f_2(\xi,\eta)$ within a window, this assumption is mathematically equivalent to assuming that they follow a zero-mean Gaussian white distribution which is independent of the distribution of disparities $d_r(\xi,\eta)$.[3]

Now we are ready to develop a statistical distribution of the intensity difference (3) between a pair of stereo images. Let us denote the right-hand side of (3) by $n_s(\xi,\eta)$. First, we compute the mean and variance of $n_s(\xi,\eta)$:

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

[1] We use $2\sigma_n^2$ in (2) as the variance of $n(x,y)$ to indicate that it accounts for noise added to both f_1 and f_2.

[2] The statistical model of (4) can be shown equivalent to assuming that $d_r(\xi,\eta)$ is generated by Brownian motion (refer to [2], [22]). More generally, we can assume $d_r(\xi,\eta)$ to be a fractal. This corresponds to choosing a different degree of $\xi^2 + \eta^2$ between $(0,1)$ in the variance in (4). Brownian motion is the simplest case in which the degree is $\frac{1}{2}$ (see Appendix A). However, our preliminary experiments have shown no noticeable advantage of using a general fractal assumption.

[3] Given no prior knowledge of a particular class of images or scenes, this assumption is justifiable on some grounds. Brownian motion is the simplest form of fractals which are often used to create natural texture patterns. In television transmission technologies, it has been known that the image intensity difference signal z follows approximately a similar exponential distribution of the form $e^{-\alpha|z|}$ where α depends on the type of the image [3] [17]. Also, except along occluding edges where intensity change and disparity change tend to occur simultaneously, intensity patterns can in general be independent of surface shapes.

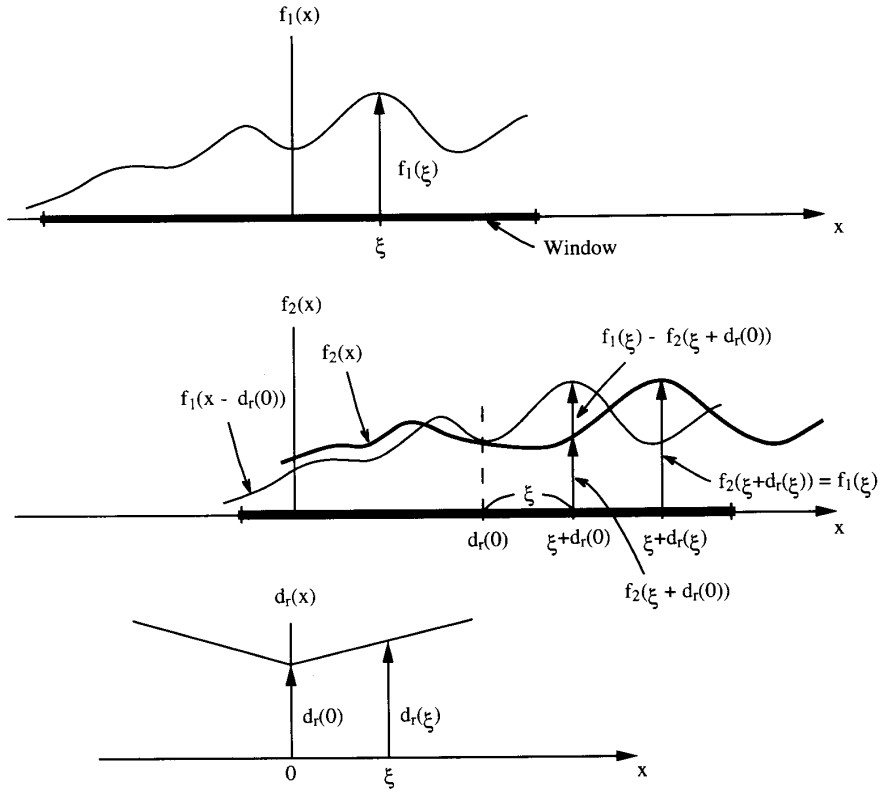


Fig. 1. Illustration of $n_s(\xi,\eta)$ in one dimension. The graph at the top shows $f_1(x)$; the middle one, $f_2(x)$ (the thicker curve) with $f_1(x)$ shifted by $d_r(0)$ (the thinner curve); the bottom one, $d_r(x)$. The regions indicated by the very thick lines on the axes indicate the region covered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

$$
\begin{aligned}
&E\left[(n_s(\xi,\eta))^2\right] \\
&= E\left[\left((d_r(\xi,\eta) - d_r(0,0))\frac{\partial}{\partial\xi} f_2(\xi + d_r(0,0),\eta)\right)^2\right] \\
&\quad + E\left[2(d_r(\xi,\eta) - d_r(0,0))\left(\frac{\partial}{\partial\xi} f_2(\xi + d_r(0,0),\eta)\right) n(\xi,\eta)\right] \\
&\quad + E\left[(n(\xi,\eta))^2\right] \\
&= E\left[(d_r(\xi,\eta) - d_r(0,0))^2\right] E\left[\left(\frac{\partial}{\partial\xi} f_2(\xi + d_r(0,0),\eta)\right)^2\right] \\
&\quad + E\left[(n(\xi,\eta))^2\right] \\
&= 2\sigma_n^2 + \alpha_f \alpha_d \sqrt{\xi^2 + \eta^2}, \qquad (6)
\end{aligned}
$$

where

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

Appendix B shows that $n_s(\xi,\eta)$ is white noise and, for mathematical convenience, its distribution can be approximated by a Gaussian distribution with the above mean and variance. That is,

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

The intuitive interpretation of (8) is as follows. Referring to Fig. 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), we can model it by zero-mean Gaussian noise whose variance (power) 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, rather than decrease, 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., the surface is frontoparallel and $\alpha_d = 0$), 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. Also, note that the noise effect of the disparity variation is amplified by a factor of α_f, that is, by the amount of the intensity variation. This is because wrong correspondences due to disparity variation affect more severely when the intensity variation is higher.

B. Models of Disparity Distribution and Local Support in Stereo

Binocular stereo matching is in general ambiguous: there are often multiple equally good matches if the matching quality is evaluated independently at each point purely by using image properties, such as area correlation, edge orientation, and slope of Laplacian zero-crossing. In order to increase the reliability of matching, all the stereo matching algorithms developed so far examine the candidate matches by calculating how much support they receive from their local neighborhood. The manner in which this support from the local neighborhood is calculated varies between algorithms and is related to fundamental assumptions the algorithms make about the scene and its surfaces. Some algorithms state such assumptions very explicitly and others rather implicitly. It is interesting and revealing to compare our statistical model of the disparity distribution ((4)) with the assumptions about local support used in other stereo algorithms.

Hakkarainen, Little, Lee and Wyatt [8] present an excellent comparison of the three most representative local support assumptions: the depth constancy assumption in Marr-Poggio stereo [12], [5], the disparity gradient limit by Pollard, Mayhew and Frisby [19], and the disparity similarity function of Prazdny [21]. The original cooperative algorithm by Marr and Poggio (MP) [12] uses a depth constancy assumption about the scene, and a match at a point looks for support from the matches in its local neighborhood which have the same disparity. Following Hakkarainen, Little, Lee and Wyatt [8], the diagram shown in Fig. 2(a) provides a graphical representation of this local support assumption. A one-dimensional case is shown for simplicity where only a neighborhood along an epipolar line is considered. The horizontal and vertical axes represent the pixel position ξ and the disparity d, respectively, relative to those of the match of concern indicated by $\square$. The thick segment on the horizontal axis on both sides of the origin indicates the region (i.e., the combinations of ξ and d) that can contribute to support the match at $\xi = 0$ for $d = 0$ (relatively speaking): that is, the neighborhood ($|\xi| \leq \xi_{max}$) whose matches have the same disparity ($d = 0$) provides support. Basically, the MP stereo assumes frontoparallel surfaces, and disparity changes are discouraged. Grimson [7] relaxes this assumption and allows neighboring points with disparities within a certain range to provide local support. Thus, the support assumption of Grimson's stereo can be represented as a rectangular region as shown in Fig. 2(b).

Pollard, Mayhew, and Frisby (PMF) [19] place a limit on the disparity gradient for acceptable matches, where a disparity gradient is defined as the ratio of the disparity difference between two points to their distance apart; the disparity gradient limit assumption means $|d/\xi| \leq g_m$. This assumption is based on the observation that the disparity gradient for correct matches is small in most cases of binocular stereo. The gradient limit constrains the relative "jaggedness" of surfaces. In its implementation, the PMF stereo computes a local support in such a way that a potential match at a point receives support only from the neighboring matches that satisfy the disparity gradient limit, while the strength of support is defined so that a closer neighborhood with a better image match gives more support [20]. Figure 2(c) shows the region in the $d - \xi$ plane which can provide support to the match at the origin. We see that the MP assumption corresponds to the case where $g_m = 0$.

Prazdny [21] argues that the major mechanism in disambiguating disparity assignments is the "coherence principle," which states that neighboring disparities, if corresponding to the same 3-D object, should be similar. Two neighboring pixels with similar disparities should support (or facilitate) each other, while pixels with dissimilar disparities should not inhibit (or interact) with each other. To incorporate this idea into a stereo algorithm, a function is needed which specifies the amount of support between neighboring points based on their disparities. Prazdny set three requirements for the function: it should be inversely proportional to the difference of disparities; more distant points should exert less influence; and the more distant the two points are, the less seriously should their disparity difference be considered. As a function which satisfies these requirements, Prazdny chose

$$s(i,j) = \frac{1}{c|i-j|\sqrt{2\pi}} e^{-\frac{|d_j - d_i|^2}{2c^2|i-j|^2}} \tag{9}$$

where $s(i,j)$ represents the amount of support that disparity d_i at pixel i receives from disparity d_j at j, $|i-j|$ is the distance between the two pixels, and c is a scaling constant. Graphically, Fig. 2(d) shows the region which exerts a support more than a certain threshold, i.e., $s(i,j) \geq s_T$. Note that $d_j - d_i$ corresponds to d and $i - j$ to ξ in our diagram.

Prazdny's similarity function (9) is exactly the same as our model of disparity distribution (4) in that the disparity difference between two pixels $d_j - d_i$ follows a zero-mean Gaussian distribution whose variance increases with their distance apart. The only difference is that in (9) the variance is proportional to the square of the distance between the pixels, instead of the distance itself as in (4). In fact, (9) is the limiting case of $H \rightarrow 1$ in (26) in Appendix A in which a general fractal surface assumption is discussed. Figure 2(e) shows the region for which $Prob(d_r(\xi) - d_r(0)) \geq P_T$. Note that in both Figs. 2(d) and (e) the support becomes stronger as the disparities become similar ($d \rightarrow 0$) and the pixels become closer ($\xi \rightarrow 0$).

Prazdny presented several computational and psychophysical arguments to justify his choice of the support function, including the relationship of the term of the exponent $|d_j - d_i|/|j - i|$ to the disparity gradient. The function represents the bias toward frontoparallel planes, but as all the diagrams in Fig. 2 show, it is a graceful mix of the distance and the disparity difference into a support score. If we view the likelihood of disparity relative to the neighboring disparity as the support score, our model is consistent with Prazdny's support function.

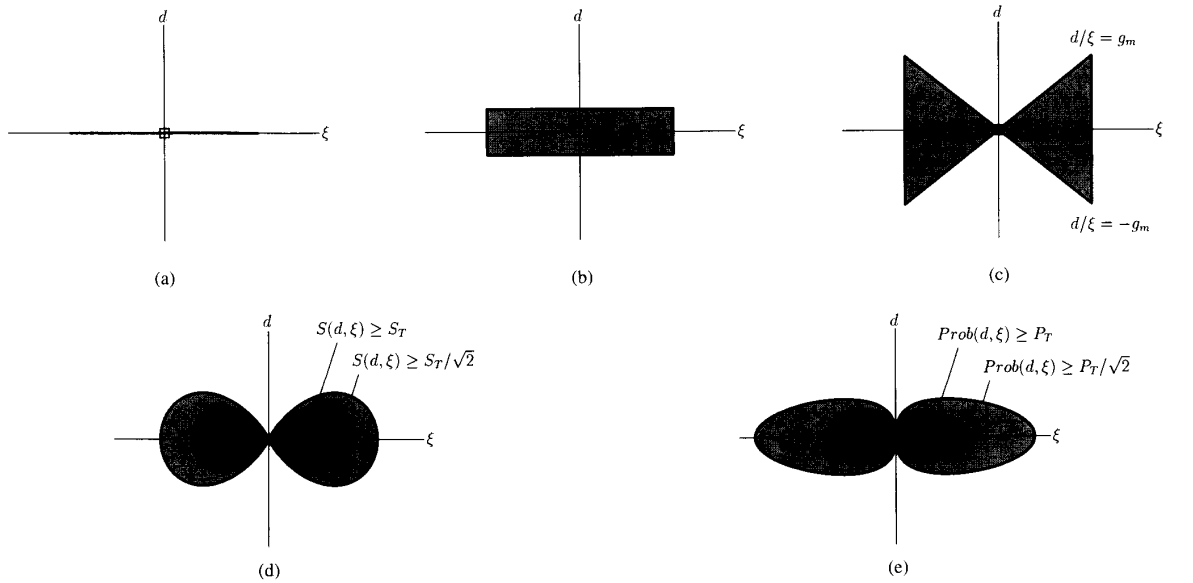


Fig. 2. Graphical representation of various local-support assumptions. Each diagram shows the local support region that provides support to the match of concern □: (a) Marr–Poggio continuity assumption; (b) Grimson's neighboring points with disparities within a certain range; (c) Pollard-Mayhew-Frisby disparity gradient limit assumption; (d) Prazdny's support in his similarity function; (e) the disparity distribution model of (4) in this paper.

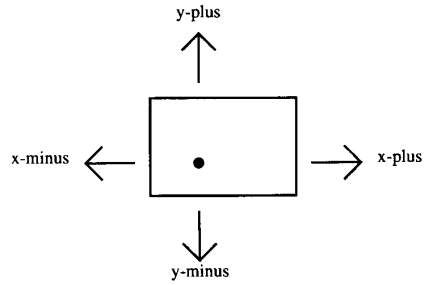


Fig. 3. Window expansion

All of the above assumptions on local support and disparity distribution tend to emphasize frontoparallel planes. As pointed out by many researchers [8], [12], [5], [7], [19], [21], however, this is not necessarily a problem. Those stereo algorithms which use such assumptions have produced acceptable results for a scene which contains slanted surfaces, curved surfaces and even disparity jumps.

It should be noted that any stereo algorithm which involves some kind of smoothing or averaging over an area does indeed assume or have a bias towards more or less frontoparallel planes, whether or not the assumption is stated explicitly. Matching by SSD calculation, for example, requires in theory (assumes implicitly) the surface to be covered by a window to have the same disparity (i.e., a frontoparallel plane) in order for it to generate an exact estimate of disparity. Otherwise, the different foreshortening occurs and the estimate becomes uncertain. The disparity distribution model of (4) allows us to use the strength of local support for evaluating that uncertainty, as opposed to using it for selecting the best match by MP, Grimson, PMF or Prazdny stereo, so that we can choose an appropriate window size which will generate the most certain estimate of disparity. The model does not necessarily limit the surface types to which the resultant stereo method will be applicable.

III. 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, the variable $n_s(\xi,\eta)$ in (8) is equal to

$$f_1(\xi,\eta) - f_2(\xi + d_0(0,0),\eta) - \Delta d \frac{\partial}{\partial \xi} f_2(\xi + d_0(0,0),\eta), \quad (10)$$

where $\Delta d = d_r(0,0) - d_0(0,0)$. Note that Δd is an incremental correction of the estimate to be made. Let us denote

$$\phi_1(\xi,\eta) = f_1(\xi,\eta) - f_2(\xi + d_0(0,0),\eta) \quad (11)$$

$$\phi_2(\xi,\eta) = \frac{\partial}{\partial \xi} f_2(\xi + d_0(0,0),\eta). \quad (12)$$

Functions ϕ_1 and ϕ_2 are the image differences and image derivatives, respectively, within a window which is placed according to the initial estimate d_0. Using these notations we can rewrite (8) as

$$n_s(\xi,\eta) = \phi_1(\xi,\eta) - \Delta d \phi_2(\xi,\eta) \sim N(0, \sigma_s^2(\xi,\eta)), \quad (13)$$

where

$$\sigma_s^2(\xi,\eta) = 2\sigma_n^2 + \alpha_f \alpha_d \sqrt{\xi^2 + \eta^2}. \quad (14)$$

Now, by sampling image values f_1 and f_2 at (ξ_i,η_j) in the window W we obtain a sample φ_{ij} of $n_s(\xi,\eta)$

$$\varphi_{ij} = n_s(\xi_i,\eta_j) = \phi_1(\xi_i,\eta_j) - \Delta d \phi_2(\xi_i,\eta_j). \quad (15)$$

From (13), the conditional probability density function of φ_{ij} given Δd is

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

Since $n_s(\xi,\eta)$ is white noise, the φ_{ij}'s are mutually independent. So the joint distribution of φ_{ij}'s for the points in the window is

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

where $\prod_{i,j\in W}$ denotes the product over the window.

The task is to estimate Δd given measurements φ_{ij}'s. Therefore, using the continuous version of Bayes' theorem we compute

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

Assuming no prior information of Δd (i.e., $p(\Delta d) =$ constant), substitution of (16) and (17) into (18) yields (see Appendix C for derivation):

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

where

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

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

where $\sum_{i,j\in W}$ denotes the summation over the window. Or, by substituting (11), (12), and (14) into (20) and (21), we obtain

$$\hat{\Delta d} = \frac{\sum_{i,j\in W} \frac{(f_1(\xi_i,\eta_j) - f_2(\xi_i + d_0(0,0),\eta_j))\frac{\partial}{\partial\xi} f_2(\xi_i + d_0(0,0),\eta_j)}{2\sigma_n^2 + \alpha_f\alpha_d\sqrt{\xi_i^2+\eta_j^2}}}{\sum_{i,j\in W} \frac{(\frac{\partial}{\partial\xi} f_2(\xi_i + d_0(0,0),\eta_j))^2}{2\sigma_n^2 + \alpha_f\alpha_d\sqrt{\xi_i^2+\eta_j^2}}} \quad (22)$$

$$\sigma_{\Delta d}^2 = \frac{1}{\sum_{i,j\in W} \frac{(\frac{\partial}{\partial\xi} f_2(\xi_i + d_0(0,0),\eta_j))^2}{2\sigma_n^2 + \alpha_f\alpha_d\sqrt{\xi_i^2+\eta_j^2}}}. \quad (23)$$

Equation (19) says that the conditional probability density function of Δd given the observed stereo image 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 (22) and (23). 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.

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

$$\hat{\alpha_d} = \frac{1}{N_w}\sum_{i,j\in W} \frac{(d_0(\xi_i,\eta_j) - d_0(0,0))^2}{\sqrt{\xi_i^2+\eta_j^2}} \quad (24)$$

$$\hat{\alpha_f} = \frac{1}{N_w}\sum_{i,j\in W} \left(\frac{\partial}{\partial\xi} f_2(\xi_i + d_0(0,0),\eta_j)\right)^2, \quad (25)$$

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

In summary, given images f_1 and f_2, a window W, and the current estimate of disparities $d_0(\xi,\eta)$ within the window, use of (22)–(25) will enable us to calculate a better estimate of disparity $d_0(0,0) + \Delta d$ at the center of the window, as well as the uncertainty of this estimation. The goal of our stereo algorithm will now become finding the disparity estimate with the lowest uncertainty.

IV. 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 present a complete description of our iterative stereo algorithm with an adaptive window.

1) Start with an initial disparity estimate[4] $d_0(x,y)$.

2) For each point (x,y), we want to choose a window that provides the estimate of disparity increment having the lowest uncertainty. For the chosen window, calculate the disparity increment by (22) 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 Fig. 3, 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 (23)–(25).

[4] As an algorithm iteratively updates disparity estimates, it requires some initial disparity estimates to start with. The initial estimates can be obtained by an existing stereo algorithm. One of the simplest is an area-based method using SSD as the matching criteria; i.e., for each pixel in the one of the images, search along the epipolar line for the matching point, which gives the minimum SSD over a small window at the pixel resolution. We used this method for the following examination in this section. Also, it is clear that this algorithm can be embedded in a coarse-to-fine control strategy to increase the range of search. In fact, for non-ambiguous stereo problems, such as those in Figs. 4, 10, and 16, we have found that by using the coarse-to-fine strategy in which we set the initial estimate as $d_0 = 0$ everywhere at the lowest-resolution level of the pyramid we can achieve basically the same results. Such a strategy, however, cannot overcome the difficulty of ambiguity in matching, such as the background part of Fig. 12; of course no other method can unless it is equipped to do so. Therefore, for Fig. 12 we used a technique of multiple-baseline stereo matching [16] to disambiguate matching in the initial estimates. While the selection of the initial estimates is important, understanding and specifying the exact conditions they must satisfy for the final success is beyond the scope of this paper.

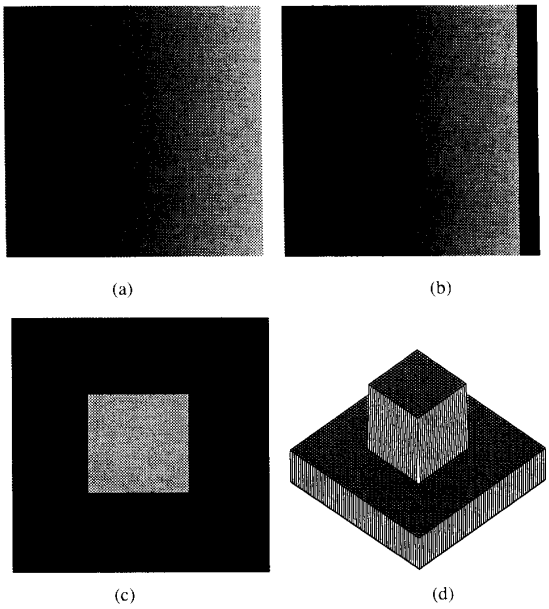


Fig. 4. Synthesized stereo images, with a ramp intensity pattern with Gaussian noise: (a) Left image; (b) Right image; (c) Disparity pattern; (d) An isometric plot of the disparity pattern.

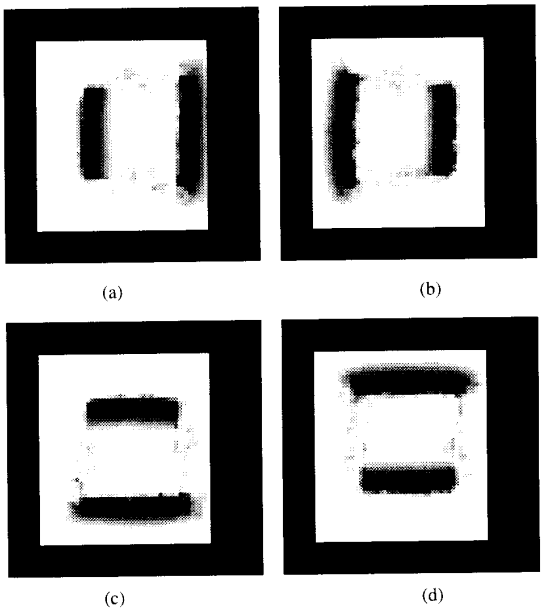


Fig. 5. Extent of window-size expansion for each direction: (a) Left (X-minus) direction; (b) Right (X-plus) direction; (c) Down (Y-minus) direction (d) Up (Y-plus) direction.

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 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.

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 4(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 Figs. 4(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. The four images in Fig. 5 show the length (increasing brightness corresponds to increasing length) by which the window has been extended in each of the four directions.[5] For example, the vertical dark stripes in Fig. 5 (a) on the right-hand side of the vertical disparity edge show that the windows for those points are not extended to the left so that the windows do not cross the disparity edge to a region of different disparity. We observe the same phenomena in the other directions. We can examine the size and shape of the selected windows at several representative positions shown in Fig. 6. The windows selected at those positions are drawn by dashed lines in Fig. 7 relative to the disparity edges drawn by solid 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 an iterative fixed-window-size SSD-based stereo method, that is, by running the same iterative algorithm except that in Step 2 of the stereo algorithm a window of predetermined size is used assuming a constant disparity over the window. We run with three window sizes, 3×3, 7×7, and 15×15. Figures 8 (a)–(c) show the result produced by fixed window sizes, and (d) 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

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

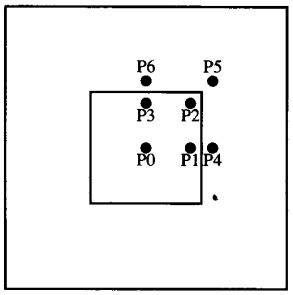


Fig. 6. Positions for which size and shape of selected windows are examined.

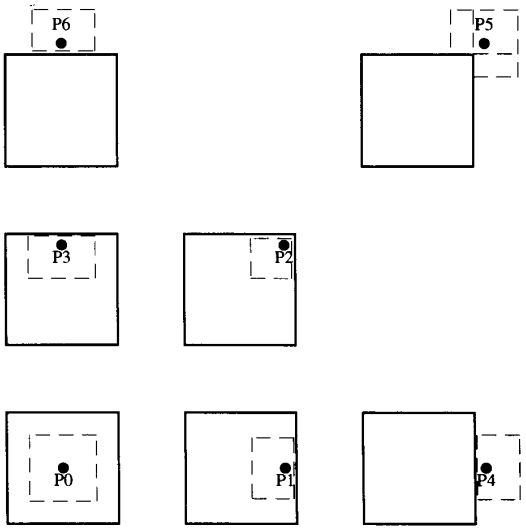


Fig. 7. Selected windows for each position

of noisy surfaces. The computed disparity by the adaptive-window algorithm shown in Fig. 8(d) 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 Fig. 9, 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 simultaneously.

Figure 10(a) and (b) show another example of synthesized test data. Figure 11 presents the computed disparity by the new method in (d), together with the results produced by fixed window sizes in (a) to (c) for comparison. As with the previous example, we clearly see better performance with the new method. The behavior of the window-size adaptation has been analyzed theoretically and tested with synthesized signals for various cases of disparity patterns including step, linear, and quadratic functions [15].

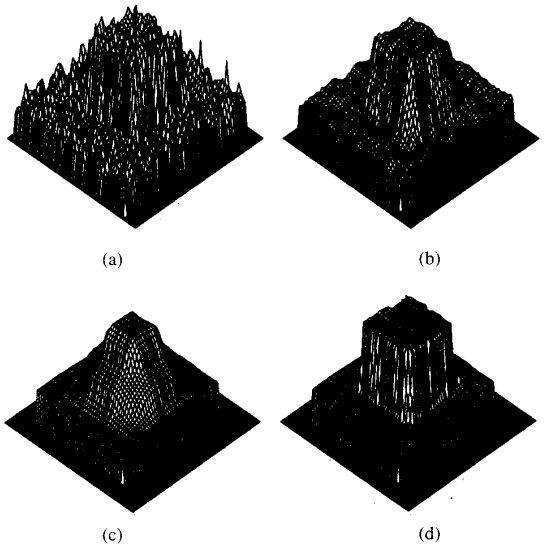
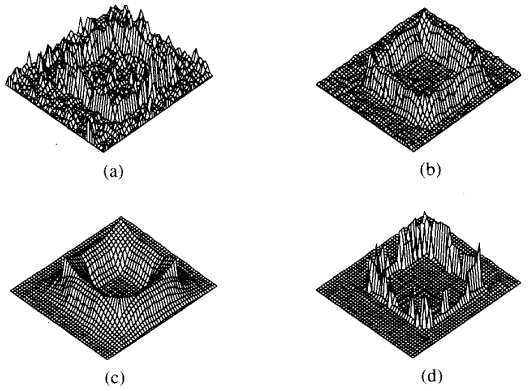


Fig. 8. Isometric plots of the computed disparity by: (a) a 3×3 window; (b) a 7×7 window; (c) a 15×15 window; (d) the adaptive window algorithm.



Window	Mean Error Value (pixel)
3×3	0.22
7×7	0.20
15×15	0.34
Adaptive Window	0.08

Fig. 9. Difference between the true disparity and the computed disparity: (a) by a 3×3 window; (b) by a 7×7 window; (c) by a 15×15 window; (d) by the adaptive window.

V. Experimental Results With Real Images

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

Figure 12 shows images of a town model that were taken by vertically-displaced cameras. The disparity, therefore, is in the vertical direction. The disparity range of this stereo pair is 4 to 14 pixels. To give an idea of the arrangement of objects in the scene, a picture taken from an oblique angle is given in Fig. 12 (c).

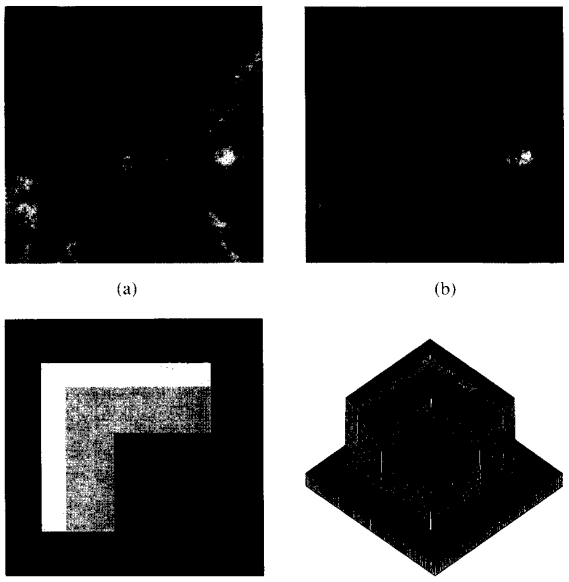


Fig. 10. Synthesized stereo images no. 2: (a) Left image; (b) Right image; (c) Disparity pattern; (d) Isometric plot of the disparity pattern shown in (c).

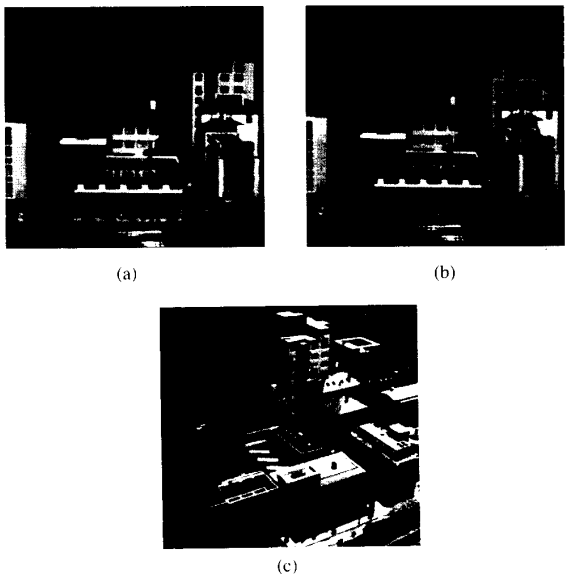


Fig. 12. "Town" stereo data set: (a) Upper image of stereo; (b) Lower image of stereo; (c) Oblique view.

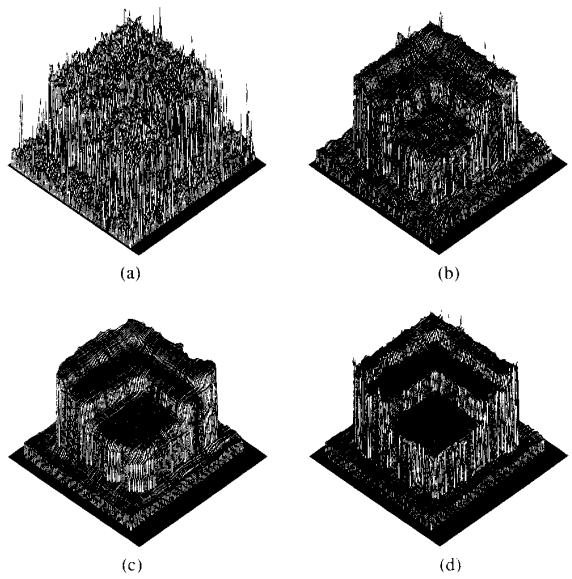
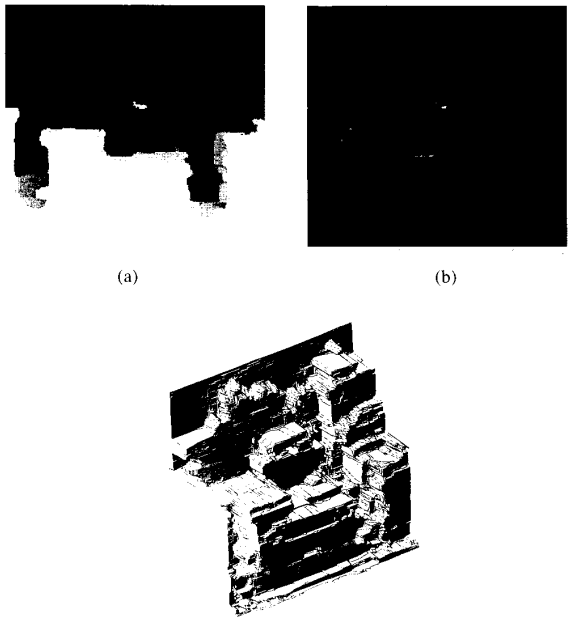


Fig. 11. Computed disparities by: (a) a fixed 3×3 window; (b) a fixed 7×7 window; (c) a fixed 15×15 window; (d) the adaptive window.

For initial disparity estimates for the experiments in this section, we have used a technique of multiple-baseline stereo matching [16] in order to remove matching ambiguities due to repetitive patterns, as shown in the top portion of Figs. 12(a) and (b). The number of iterations in step 3 of the algorithm description was set to 5. Figure 13(a) shows the disparity map computed by the adaptive window algorithm. In addition, the uncertainty estimate computed by the algorithm is shown in Fig. 13(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 Fig. 13(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 isometric plot of the disparity map is shown in Fig. 13(c), which roughly corresponds to the viewing angle of Fig. 12(c). We can see that each building wall has a smooth surface and yet is clearly separated from others, and the shape of the distant bridge (on the left) is recovered. For comparison, the resultant isometric plots of the disparity maps with fixed window sizes are shown in Figs. 14(a) 3×3, (b) 7×7, and (c) 15×15. We observe noisy surface reconstraction by a small window and over-smoothing of disparity edges by a large window.

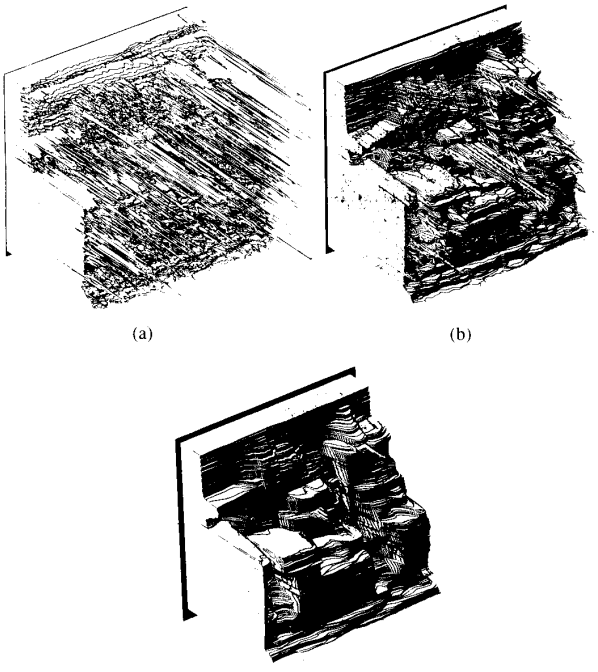
Figure 15 presents perspective views of the recovered scene by texture mapping the original intensity image on the constructed disparity map shown in Fig. 13 and generating views from new positions which are outside of the original stereo views. They can give an idea of the quality of reconstruction. This stereo data set is the same one used in [13]. 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 16 (a) and (b) show another set of real stereo images which are top views of a coal mine model. The disparity range is 30 to 40 pixels. Figures 17 (a) and (c) show the isometric plots of the computed disparity. For comparison, actual pictures of the model taken from roughly the same angles are given in figures 17 (b) and (d). The shapes of buildings, a $\wedge$-shaped roof, a water tank on the roof, and flat ground have been recovered without blurring the edges.



(a) (b) (c)

Fig. 13. Disparity and uncertainty computed by the adaptive window algorithm for the "town" stereo data: (a) Disparity map; (b) Uncertainty; (c) Isometric plot of the disparity map.



(a) (b) (c)

Fig. 14. Isometric plots of the disparity maps computed by fixed-size windows: (a) 3 × 3; (b) 7 × 7; (c) 15 × 15.

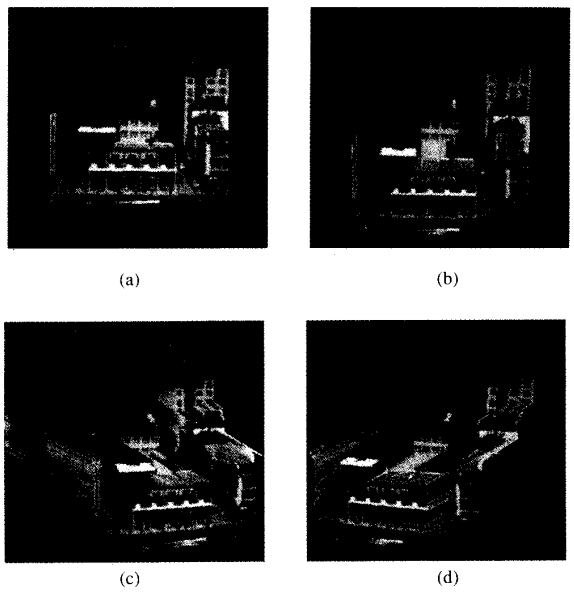
VI. Conclusion

In this paper, we have presented an iterative stereo matching algorithm using an adaptive window. The algorithm selects a window adaptively for each pixel so 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 locally adaptive window. The algorithm starts with initial estimates of disparity, and iteratively updates them by this adaptive window method.

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.



(a) (b) (c) (d)

Fig. 15. Perspective views of the recovered scene: (a) from the original camera position; (b) from an upper position; (c) from an upper left position; (d) from an upper right position.

(a) (b)

Fig. 16. "Coal mine" stereo data set: (a) Lower image; (b) Upper image.

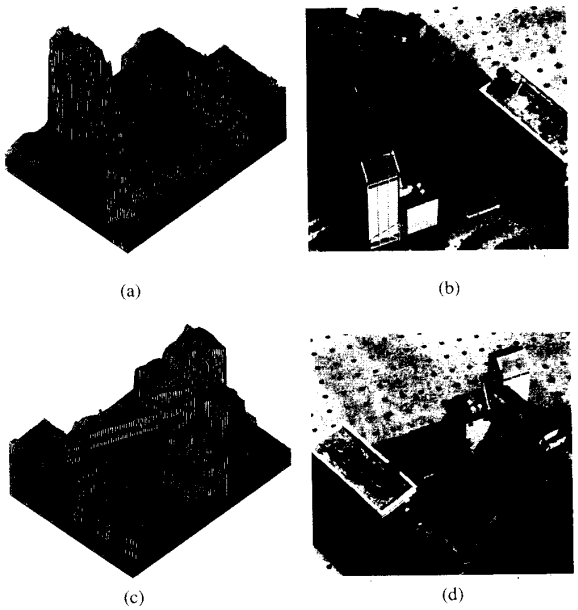


Fig. 17. Isometric plots of the computed disparity map and their corresponding actual view: (a), (b) Isometric plot and corresponding view from the lower left corner; (c), (d) Isometric plot and corresponding view from the upper right corner.

APPENDIX A
ASSUMING FRACTAL FOR $d_r(\xi,\eta)$

Here we assume $d_r(\xi,\eta)$ to be fractal, then instead of (4), we have

$$d_r(\xi,\eta) - d_r(0,0) \sim N\big(0, \alpha_d(\xi^2+\eta^2)^H\big), \tag{26}$$

where the parameter H has a value $0 < H < 1$. When $H = \frac{1}{2}$, this equation represents the case that $d_r(\xi,\eta)$ is Brownian motion.

Then, instead of the final equations (22) and (23), we get

$$\hat{\Delta d} = \frac{\sum_{i,j\in W} \frac{(f_1(\xi_i,\eta_j) - f_2(\xi_i + d_0(0,0),\eta_j))\frac{\partial}{\partial\xi} f_2(\xi_i + d_0(0,0),\eta_j)}{2\sigma_n^2 + \alpha_f \alpha_d (\xi_i^2+\eta_j^2)^H}}{\sum_{i,j\in W} \frac{(\frac{\partial}{\partial\xi} f_2(\xi_i + d_0(0,0),\eta_j))^2}{2\sigma_n^2 + \alpha_f \alpha_d (\xi_i^2+\eta_j^2)^H}}$$

$$\sigma^2_{\Delta d} = \frac{1}{\sum_{i,j\in W} \frac{(\frac{\partial}{\partial\xi} f_2(\xi_i + d_0(0,0),\eta_j))^2}{2\sigma_n^2 + \alpha_f \alpha_d (\xi_i^2+\eta_j^2)^H}}.$$

Furthermore, instead of (24), we obtain:

$$\hat{\alpha_d} = \frac{1}{N_w} \sum_{i,j\in W} \frac{(d_0(\xi_i,\eta_j) - d_0(0,0))^2}{(\xi_i^2+\eta_j^2)^H}.$$

APPENDIX B
APPROXIMATING DISTRIBUTION OF $n_s(\xi,\eta)$

We will examine the statistical properties of $n_s(\xi,\eta)$, which is the right-hand side of (1)

$$n_s(\xi,\eta) = (d_r(\xi,\eta) - d_r(0,0)) \cdot \frac{\partial}{\partial\xi} f_2(\xi + d_r(0.0).\eta) + n(\xi.\eta).$$

We see that $n_s(\xi,\eta)$ has the form of

$$xy + n,$$

where

$$x = d_r(\xi,\eta) - d_r(0,0),$$
$$y = \frac{\partial}{\partial\xi} f_2(\xi + d_r(0,0),\eta),$$
$$n = n(\xi,\eta).$$

Our assumptions are that: x is a zero-mean Gaussian noise; y and n are both zero-mean Gaussian white noise; and x, y, and n are statistically independent.

Let $p_x(x)$, σ_x^2, and $R_x(\tau)$ denote the density function, variance, and autocorrelation function of x, respectively.

$$p_x(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{x^2}{2\sigma_x^2}}.$$

We define notations for y and n in the same manner. Since y is white, we have

$$R_y(\tau) = a\delta(\tau),$$

where, $\delta(\tau)$ is the delta function and a is a constant.

First, let us examine the properties of random variable z which is product of x and y

$$z = xy.$$

Since x and y are independent, the autocorrelation function of z is given by (see [4]):

$$R_z(\tau) = R_x(\tau)R_y(\tau) = b\delta(\tau),$$

where b is a constant. Therefore, z is also white.

The density function $p_z(z)$ can be calculated as

$$\begin{aligned} p_z(z) &= \int_{-\infty}^{\infty} \frac{1}{|x|} p_x(x) p_y\Big(\frac{z}{x}\Big) dx \\ &= \frac{1}{\pi\sigma_x\sigma_y} \int_0^{\infty} \frac{1}{x} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{z^2}{2\sigma_y^2 x^2}\right) dx \\ &= \frac{1}{\pi\sigma_x\sigma_y} K_0\Big(\frac{|z|}{\sigma_x\sigma_y}\Big), \end{aligned}$$

where $K_0(z)$ is the modified Bessel function of order 0

$$K_0(z) = \frac{1}{2}\int_0^{\infty} x^{-1} \exp\left(-x - \frac{z^2}{4x}\right) dx.$$

The thick curve in Fig. 18 shows this density function. $p_z(z)$ is a monomodal distribution which is symmetrical about the mode at $z = 0$. For simplicity, it is reasonable to approximate the distribution by a Gaussian distribution that has the same mean and variance as $p_z(z)$, which are

$$\begin{aligned} E[z] &= E[x]E[y] \\ &= 0 \\ E[(z - E[z])^2] &= E[(xy)^2] = E[x^2]E[y^2] \\ &= \sigma_x^2\sigma_y^2. \end{aligned}$$

$$
\begin{aligned}
p(\Delta d|\varphi_{ij}(i,j\in W)) &= \frac{\prod_{i,j\in W}\frac{1}{\sqrt{2\pi}\sigma_s(\xi_i,\eta_j)}\exp\left(-\frac{(\phi_1(\xi_i,\eta_j)-\Delta d\phi_2(\xi_i,\eta_j))^2}{2\sigma_s^2(\xi_i,\eta_j)}\right)}{\int_{-\infty}^{\infty}\prod_{i,j\in W}\frac{1}{\sqrt{2\pi}\sigma_s(\xi_i,\eta_j)}\exp\left(-\frac{(\phi_1(\xi_i,\eta_j)-\Delta d\phi_2(\xi_i,\eta_j))^2}{2\sigma_s^2(\xi_i,\eta_j)}\right)d(\Delta d)} \\
&= \frac{\exp\left(\sum_{i,j\in W}\left(-\frac{(\phi_1(\xi_i,\eta_j)-\Delta d\phi_2(\xi_i,\eta_j))^2}{2\sigma_s^2(\xi_i,\eta_j)}\right)\right)}{\int_{-\infty}^{\infty}\exp\left(\sum_{i,j\in W}\left(-\frac{(\phi_1(\xi_i,\eta_j)-\Delta d\phi_2(\xi_i,\eta_j))^2}{2\sigma_s^2(\xi_i,\eta_j)}\right)\right)d(\Delta d)} \\
&= \frac{\exp\left(-\frac{\sum_{i,j\in W}\phi_2(\xi_i,\eta_j)^2}{2\sigma_s^2(\xi_i,\eta_j)}\left(\Delta d-\frac{\sum_{i,j\in W}(\phi_1(\xi_i,\eta_j)\phi_2(\xi_i,\eta_j)/\sigma_s^2(\xi_i,\eta_j))}{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2}\right)^2\right)}{\int_{-\infty}^{\infty}\exp\left(-\frac{\sum_{i,j\in W}\phi_2(\xi_i,\eta_j)^2}{2\sigma_s^2(\xi_i,\eta_j)}\left(\Delta d-\frac{\sum_{i,j\in W}(\phi_1(\xi_i,\eta_j)\phi_2(\xi_i,\eta_j)/\sigma_s^2(\xi_i,\eta_j))}{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2}\right)^2\right)d(\Delta d)} \\
&= \sqrt{\frac{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2}{2\pi}}\exp\left(-\frac{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2}{2}\right. \\
&\qquad \left.\left(\Delta d-\frac{\sum_{i,j\in W}(\phi_1(\xi_i,\eta_j)\phi_2(\xi_i,\eta_j)/\sigma_s^2(\xi_i,\eta_j))}{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2}\right)^2\right)
\end{aligned}
$$

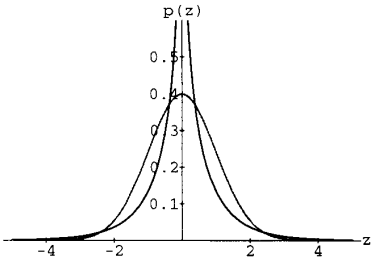


Fig. 18. Probability density functions, $\frac{1}{\pi\sigma_x\sigma_y}K_0\left(\frac{|z|}{\sigma_x\sigma_y}\right)$ and $N(0,\sigma_x^2\sigma_y^2)$. The horizontal axis is normalized; i.e., $z' = \frac{z}{\sigma_x\sigma_y}$

The faint curve in Fig. 18 shows the zero-mean Gaussian distribution $N(0,\sigma_x^2\sigma_y^2)$.

Once we approximate the white noise $z = xy$ by a Gaussian distribution, it is straightforward to see that $z + n$ is also a Gaussian white noise and to calculate its mean and variance, since it is sum of two Gaussian white noises. Hence, (8).

APPENDIX C
DERIVATION OF (19)–(21)

Substituting (16) and (17) into (18), we get the equation found at the top of the page, where $\sum_{i,j\in W}$ denotes the summation over the window. From this equation, we can see that $p(\Delta d|\varphi_{ij}(i,j\in W))$ becomes a Gaussian probability density function. Its mean value $\hat{\Delta d}$ and variance $\sigma_{\Delta d}^2$ are

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

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

Hence, (19) to (21).

REFERENCES

[1] S. T. Barnard and M. A. Fischler, "Stereo vision," in *Encyclopedia of Artificial Intelligence*. New York: John Wiley, 1987, pp. 1083–1090.

[2] B. B. Mandelbrot and B. J. Van Ness, "Fractional brownian motion, fractional noises and applications," *SIAM*, vol. 10, no. 4, pp. 422–438, 1968.

[3] J. C. Candy, M. A. Franke, B. G. Haskell, and F. W. Mounts, "Transmitting television as clusters of frame-to-frame differences," *Bell Syst. Tech. J.*, vol. 50, no. 6, pp. 1889–1919, 1971.

[4] F. deCoulon, *Signal Theory and Processing*. Norwood, MA: Artech House, Inc., 1986.

[5] M. Drumheller and T. Poggio, "On parallel stereo," in *Proc. Int. Conf. Robotics and Automat.*, 1986, pp. 1439–1448.

[6] W. Forstner and A. Pertl, *Photogrammetric Standard Methods and Digital Image Matching Techniques for High Precision Surface Measurements*. New York: Elsevier Science Publishers B.V., 1986, pp. 57–72.

[7] W. E. L. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 7, no. 1, pp. 17–34, Jan. 1985. (The shape of the support region is due to personal communication.)

[8] J. M. Hakkarainen, J. J. Little, H. Lee, Jr. and J. L. Wyatt, "Interaction of algorithm and implementation for analog VLSI stereo vision," in *SPIE Visual Inform. Processing: From Neurons to Chips*, pp. 173–184, 1991.

[9] M. J. Hannah, "A system for digital stereo image matching," *Photogrammetric Engineering and Remote Sensing*, vol. 55, no. 12, pp. 1765–1770, Dec. 1989.

[10] W. Hoff and N. Ahuja, "Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 2, 1989.

[11] M. D. Levine, D. A. O'Handley, and G. M. Yagi, "Computer determination of depth maps," *Comput. Graphics and Image Processing*, vol. 2, no. 4, pp. 131–150, 1973.

[12] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, pp. 283–287, Oct. 1976.

[13] L. Matthies, R. Szeliski, and T. Kanade, "Kalman filter-based algorithms for estimating depth from image sequences," *Int. J. Comput. Vision*, vol. 3, pp. 209–236, 1989.

[14] K. Mori, M. Kidode, and H. Asada, "An iterative prediction and correction method for automatic sterocomparison," *Comput. Graphics and Image Processing*, vol. 2, pp. 393–401, 1973.

[15] M. Okutomi and T. Kanade, "A locally adaptive window for signal matching," in *Proc. Int. Conf. Comput. Vision*, Dec. 1990. Also appeared in CMU Tech. Rep. CMU-CS-90-178, 1990.

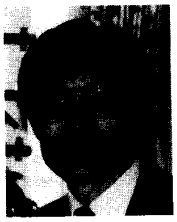
[16] ——, "A multiple-baseline stereo," in *Proc. Comput. Vision Pattern Recognit.*, June 1991, pp. 63–69. Also in *IEEE Trans. Pattern Anal.*

Machne Intell., vol. 15, no. 4, pp. 353–363, 1993.
[17] J. B. O'Neal, "Predictive quantizing systems for the transmission of television signals," *Bell Syst. Tech. J.*, vol. 45, no. 5, pp. 689–722, 1966.
[18] D. J. Panton, "A flexible approach to digital stereo mapping," *Photogram. Eng. Remote Sensing*, vol. 44, no. 12, pp. 1499–1512, Dec. 1978.
[19] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby, "Pmf: A stereo correspondence algorithm using a disparity gradient limit," *Perception*, vol. 14, pp. 449–470, 1985.
[20] S. B. Pollard, J. Porrill, J. E. W. Mayhew, and J. P. Frisby, "Disparity gradient, Lipschitz continuity, and computing binocular correspondences," in *Robotics Research, The Third International Symposium*, O. Faugeras and G. Giralt Eds. Cambridge MA: The MIT Press, 1986, pp. 19–26.
[21] K. Prazdny, "Detection of binocular disparities," *Biological Cybern.*, vol. 52, pp. 93–99, 1985.
[22] R. F. Voss, "Fractals in nature," in *Course note on FRACTALS: Introduction, Basics, and Perspectives*, 1987.
[23] G. A. Wood, "Realities of automatic correlation problem," *Photogrammetric Engineering and Remote Sensing*, vol. 49, pp. 537–538, Apr. 1983.

Takeo Kanade (M'80–SM'88–F'92) received the Doctoral degree in electrical engineering from Kyoto University, Japan, in 1974.

After holding a faculty position at the Department of Information Science, Kyoto University, he joined Carnegie Mellon University in 1980, where he is currently the U.A. Helen Whitaker Professor of Computer Science and the Director of the Robotics Institute.

Dr. Kanade has made technical contributions in multiple areas of robotics: vision, manipulators, autonomous mobile robots, and sensors. He has written more than 150 technical papers and reports in these areas. He has been the principal investigator of several major vision and robotics projects at Carnegie Mellon. In the area of education, he was a founding chair person of CMU's Robotics Ph.D. Program, probably the first of its kind. Dr. Kanade is a Founding Fellow of the American Association of Artificial Intelligence, and the founding editor of the *International Journal of Computer Vision*. He has served for many government, industry, and university advisory or consultant committees, including Aeronautics and Space Engineering Board (ASEB) of the National Research Council, NASA's Advanced Technology Advisory Committee (Congressionally mandate committee) and the Advisory Board of Canadian Institute for Advanced Research.



Masatoshi Okutomi received the B.E. degree in mathematical engineering and information physics from the University of Tokyo, Japan, in 1981, and the M.E. degree in control engineering in 1983, and the Ph.D. degree for his research on stereo vision, both from the Tokyo Institute of Technology (TIT), Japan.

In 1983, he joined the Canon Research Center in Tokyo. From 1987 to 1990, he was a visiting research scientist with the School of Computer Science at Carnegie Mellon University, Pittsburgh, PA. In 1994, he became an Associate Professor of the Department of Control and Systems Engineering at TIT. His current research interests include both theoretical aspects of computer vision and its applications, such as robotics and human–machine interaction.