ELSEVIER

Artificial Intelligence 78 (1995) 121-145

Artificial Intelligence

Multilevel enhancement and detection of stereo disparity surfaces

Yibing Yang *, Alan L. Yuille

Division of Applied Sciences, Harvard University, Cambridge, MA 02138, USA

Received September 1993; revised December 1994

Abstract

The problem of stereo vision has been of increasing interest to the computer vision community over the past decade. This paper presents a new computational framework for matching a pair of stereo images arising from viewing the same object from two different positions. In contrast to previous work, this approach formulates the matching problem as detection of a "bright", coherent disparity surface in a 3D image called the *spatio-disparity space* (SDS) image. The SDS images represents the goodness of each and every possible match.

A nonlinear filter is proposed for enhancing the disparity surface in the SDS image and for suppressing the noise. This filter is used to construct a hyperpyramid representation of the SDS image. Then the disparity surface is detected using a coarse-to-fine control structure. The proposed method is robust to photometric and geometric distortions in the stereo images, and has a number of computational advantages. It produces good results for complex scenes.

1. Introduction

1.1. Stereo matching and feature detection

When viewed from two different perspectives, the same object will give rise to a pair of different images. The 3D shape and location of the object can be recovered by fusing the stereo pair. The human ability of stereo vision was first observed by Wheatstone [32], and the underlying mechanism has been investigated from the computational standpoint [4, 14, 18]. Stereo is a useful method for machine perception. The most important applications are cartography [8] and robot vision [20, 21].

* Corresponding author. Current address: Polaroid Co., 750M-3J, Cambridge, MA 02139, USA. E-mail: yangy@polaroid.com.

SSDI 0004-3702(95)00028-3

In recovering 3D shape using stereo, the most difficult task is *image matching*. Given a point in the left image, the problem is to find its corresponding point in the right image. (Thanks to the *epipolar* line constraint, we may assume that corresponding points lie on the same horizontal raster line.) Stereo matching is made possible by two facts: (1) Neighboring image points tend to have roughly the same disparity. Based on this fact, a geometric constraint may be introduced. (2) Corresponding points in the two images, originating from the same point on the object, appear similar; stereo matching would be impossible otherwise. Based on this fact, a photometric constraint may be formulated.

Many stereo algorithms have been proposed. As a common feature, existing approaches attempt to *match* a left-image point to a right-image point, based on the assumption that the disparity is locally constant [9-11, 13, 16, 17, 19, 24, 26, 27] or more generically, piecewise smooth [2, 3, 5, 22, 25, 31, 33].

This paper presents a new approach to stereo matching. We convert the matching problem to a *feature detection* problem. To match a pair of 1D stereo images $I^l = I^l(x)$ and $I^r = I^r(x)$, we may assign a numerical value $\Xi(x,u)$ to measure the photometric similarity between the point x in I^l and the point $(x-u)$ in I^r. By computing $\Xi(x,u)$ for every possible match (x,u) we get an *spatio-disparity space* image (SDS image) $\Xi = \Xi(x,u)$. As such the task becomes the detection the "bright" disparity curve in the SDS image $\Xi = \Xi(x,u)$. In the case of matching two conventional 2D stereo images, the problem is to detect the "bright" disparity surface in the 3D SDS image $\Xi = \Xi(x,y,u)$.

Consider the 1D case where the SDS image is $\Xi = \Xi(x,u)$. We want a filter which can enhance the disparity curve and simultaneously suppress the noise. The disparity curve has a number of photometric/geometric properties that makes it different from conventional curves such as fingerprints. The disparity curve sensitive filter should be able to take advantage of these properties. Conventional curve detectors prove ineffective in this respect.

We will present a new filter which is sensitive to the disparity surface/curve in the SDS image. The filter is quite simple, and can be synthesized from two conventional filters—a morphological filter and a linear filter. Unlike conventional curve detectors, the filter is nonlinear; the nonlinearity enables it to accommodate disparity surface undulations in an effective and efficient way.

1.2. Hierarchical methods

Hierarchical methods based on pyramid data structures have been widely used in computer vision [28, 29], and in particular, in stereo matching. Traditional hierarchical methods for stereo matching are based on the multiresolution representations [6] of the stereo images and thus can be characterized as *multiresolution methods* [1, 7, 9, 10, 17, 19, 27, 31]. Fig. 1 illustrates the multiresolution paradigm. The first step is to construct a pair of pyramids $\{I_k^l\}$ and $\{I_k^r\}$ for the input stereo images I^l and I^r. At the kth level, I_k^l and I_k^l are respectively reduced-resolution versions of I^l and I^r. Then matching is performed in a coarse-to-fine manner. At the kth level, the problem is to match I_k^l and I_k^r. The belief is that by smoothing the stereo images matching is simplified.

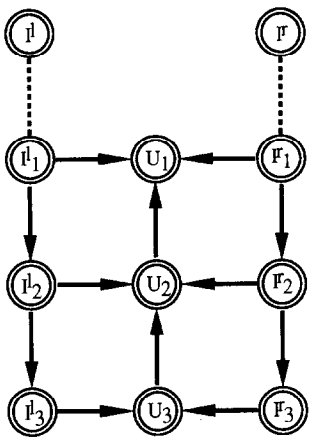

Fig. 1. The multiresolution paradigm for disparity detection.

The multiresolution scheme, however, is counterproductive for some types of scenes; by smoothing the input images stereo information in textured regions is also filtered out. From the psychophysical perspective, smoothed stereo pairs are not necessarily easier to fuse; instead, dense texture is helpful for stereo fusion. Moreover, multiresolution methods often have difficulty coping with sloping surfaces in the scene. For example, stereo images of sloping Lambertian surfaces have the regularity that corresponding image points have the same intensity. But in the smoothed images, no photometric regularity remains, unless the viewed surface is fronto-parallel.

In this paper we will propose a new hierarchical method for stereo matching called the *multilevel method.* As stated above, stereo matching amounts to detecting the disparity surface in the SDS image. The multilevel method is based on a hyperpyramid representation $\{\Xi_k\}$ of the SDS image $\Xi = \Xi(x, y, u)$. As illustrated in Fig. 2, the hyperpyramid is recursively constructed; Ξ_{k+1} is a filtered and reduced-size version of Ξ_k, where the filter is the surface-sensitive filter mentioned above; Ξ_{k+1} is not a reduced-resolution version of Ξ_k. We can envision the construction of the SDS image hyperpyramid as the integration of stereo information in I^l and I^r over successively-larger image regions. In integrating this stereo information we do not assume disparity constancy in the image regions. Once the hyperpyramid is constructed, we recursively compute the disparity surfaces $\{U_k\}$ in a coarse-to-fine manner.

2. The SDS image

2.1. Spatio-disparity space

At an image position (x, y), the disparity value u can be any real number. We call the xyu-space the *spatio-disparity space* (SDS) which consists of all possible matches between points in the left and right images. Any point (x, y, u) in the SDS represents a match.

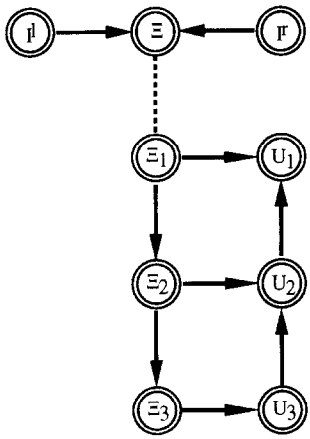

Fig. 2. The multilevel paradigm for disparity detection.

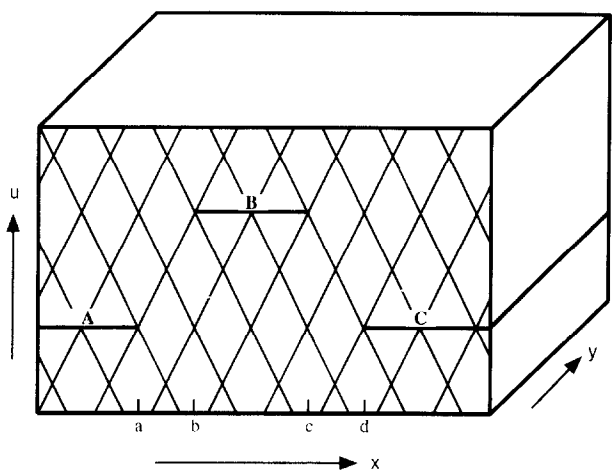

Fig. 3. Lines of sight in the cyclopean SDS.

We want to investigate the disparity surface $U(x, y)$ in the context of the SDS. The SDS can be defined with respect to the left image or the right image or the cyclopean image. In the cyclopean SDS, a point (x, y, u) represents the match between point $(x + u/2, y)$ in the left image, and point $(x - u/2, y)$ in the right image.

Fig. 3 shows the *lines of sight* for the left and right views in the cyclopean SDS. (A line of sight is defined by an image point and the corresponding focal point.) In the cyclopean SDS the right lines of sight have a slope of 2 and the left lines of sight have a slope of -2. Of course, the cyclopean lines of sight are parallel to the u-axis.

An important constraint regarding the disparity surface is the uniqueness constraint: A left line of sight or a right line of sight can intersect with the disparity surface at no more than one point. Fig. 3 shows a profile of a valid disparity surface in the xu-plane.

The region between a and b is visible only in the left image and, on the other hand, the region between c and d is visible only in the right image.

2.2. SDS image

A point in the SDS represents a match, and the goodness of the match can be evaluated by using an appropriate similarity measure. By assigning a numerical value, or intensity, to each SDS point we get an SDS image.

Given a pair of stereo images $I^l = I^l(x,y)$, $I^r = I^r(x,y)$, the SDS image can be defined by using the sum-of-squared-difference measure. Let

$$\theta_l(\xi,\eta) = I^l(x+u/2+\xi, y+\eta),$$
$$\theta_r(\xi,\eta) = I^r(x-u/2+\xi, y+\eta).$$

Then the intensity of an SDS point (x,y,u) is defined as

$$\Xi(x,y,u) = -\sum_{\xi,\eta} w(\xi,\eta)\big(\theta_l(\xi,\eta) - \theta_r(\xi,\eta)\big)\big)^2, \tag{1}$$

where w is a window function which has the following properties: (1) $w(\cdot,\cdot)$ is *separable*: $w = w(\xi,\eta) = w(\xi)w(\eta)$; (2) $w(\cdot)$ is *symmetric* and *nonnegative-definite*; (3) $w(\cdot,\cdot)$ is *normalized*: $\sum_{\xi,\eta} = 1$.

A more desirable similarity measure is *normalized correlation*. Based on this measure, the SDS image intensity can be expressed as:

$$\Xi(x,y,u) = \frac{\sum_{\xi,\eta} w(\xi,\eta)\theta_l(\xi,\eta)\theta_r(\xi,\eta)}{\sqrt{\sum_{\xi,\eta} w(\xi,\eta)(\theta_l(\xi,\eta))^2 \sum_{\xi,\eta} w(\xi,\eta)(\theta_r(\xi,\eta))^2}}, \tag{2}$$

where $0 \leqslant \Xi(x,y,u) \leqslant \Xi_{\max} \equiv 1$. This measure has the advantage of being robust to the photometric distortions in the stereo pair arising from the change of viewing position. The results reported in this paper were obtained using this measure.

2.2.1. Problem of ambiguity

In the SDS image a point on the disparity surface will appear "bright", because it represents a correct match. Ideally the disparity surface in the SDS image can be detected by maximum-picking. The expressions for the cyclopean disparity surface, the left disparity surface, and the right disparity surface are

$$U(x,y) = \arg\max_u \Xi(x,y,u), \tag{3}$$

$$U^l(x,y) = \arg\max_u \Xi(x-u/2,y,u), \tag{4}$$

$$U^r(x,y) = \arg\max_u \Xi(x+u/2,y,u), \tag{5}$$

respectively. Note that in the SDS image the left (or right) view lines of sight have a slope of -2 (or 2).

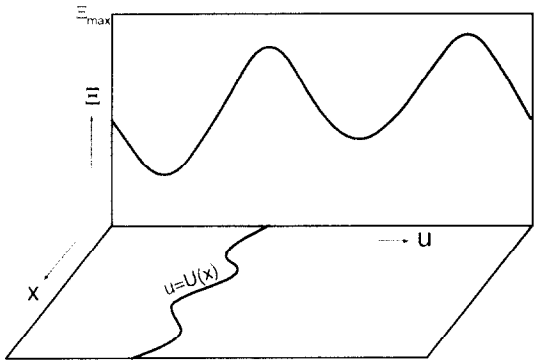

Fig. 4. Problem of local ambiguity.

However, this scheme is impractical; along a line of sight, the image intensity $\Xi = \Xi(u)$ may have multiple peaks and the true peak may be overridden by a spurious peak (see Fig. 4). In other words, the SDS image is ambiguous. Ambiguities come mainly from the following two sources:

- *Photometric source.* Ambiguities may result from image noise in the stereo pair introduced in the imaging process. In addition, lack of intensity variations and the existence of repetitive texture patterns in the stereo images all contribute to ambiguities.
- *Geometric source.* In computing the SDS image, geometric distortions are completely neglected—all image points in the window are assumed to have the same disparity.

The effect of the photometric source can be reduced by using a large window. By increasing the window size, more image information can be gathered for disambiguation, and the effect of image noise can be suppressed.

From the geometric point of view, however, a large window is undesirable. As the size of the window increases, geometric distortions grow. This problem has been addressed in [15,23]. Indeed, only in a small window is disparity invariance justified. Should a large window be used, a correct match may yield an SDS image intensity value much smaller than $\Xi_{\max}$. In short, increasing the window size does not necessarily lead to less ambiguities.

2.3. Computation of the SDS image

One may want to compute the SDS image directly from the definition. That is, to evaluate $\Xi(x, y, u)$ for each and every (x, y, u). But this brute force approach is computationally expensive. In this subsection we will present a fast algorithms for computing the SDS image.

The SDS image defined by (2) can be constructed using one-dimensional convolution operations. Before proceeding to the construction algorithm, we first introduce a few terms.

- *Local auto-correlation* of a two-dimensional image $I = I(x,y)$ is the convolution $\Phi(x,y) = (I(x,y))^2 * w(x,y)$, where $*$ denotes convolution, and "local" is specified by the window function w.
- *Local cross-correlation* of two images $I^l(x,y)$, $I^r(x,y)$ is similarly defined: $\Phi_{lr}(x,y) = \left(I^l(x,y)I^r(x,y)\right) * w(x,y)$.

The SDS image defined by (2) can expressed as a local cross-correlation divided by two local auto-correlations:

$$\Xi(x,y,u) = \frac{\Phi_{lr}(x,y;u)}{\Phi_l(x,y;u)\Phi_r(x,y;u)}, \tag{6}$$

where

- $\Phi_{lr}(x,y;u)$ is the local cross-correlation of the shifted left and right images:

$$\Phi_{lr}(x,y;u) = \left(I^l(x+u/2,y)I^r(x-u/2,y)\right) * w(x,y), \tag{7}$$

- $\Phi_l(x,y;u)$, $\Phi_r(x,y;u)$ are square roots of the local auto-correlations of the shifted left and right images respectively:

$$\Phi_l(x,y;u) = \sqrt{\left(I^l(x+u/2,y)\right)^2 * w(x,y)}, \tag{8}$$

$$\Phi_r(x,y;u) = \sqrt{\left(I^r(x-u/2,y)\right)^2 * w(x,y)}. \tag{9}$$

Thus for $u = u_1, u_2, u_3, \ldots$, we may compute a sequence of image slices $\Xi(\cdot,\cdot;u_1)$, $\Xi(\cdot,\cdot;u_2)$, $\Xi(\cdot,\cdot;u_3), \ldots$. Each image slice is computed from $\Phi_l(\cdot,\cdot;u)$, $\Phi_r(\cdot,\cdot;u)$ as well as $\Phi_{lr}(\cdot,\cdot;u)$.

The point is that we do not have to compute $\Phi_l(\cdot,\cdot;u)$, $\Phi_r(\cdot,\cdot;u)$ from (8) and (9) for every u. We observe that (8) and (9) can respectively be written as

$$\Phi_l(x,y;u) = \Phi_l(x-u/2,y), \tag{10}$$

$$\Phi_r(x,y;u) = \Phi_r(x+u/2,y), \tag{11}$$

where

$$\Phi_l(x,y) = \sqrt{\left(I^l(x,y)\right)^2 * w(x,y)}, \tag{12}$$

$$\Phi_r(x,y) = \sqrt{\left(I^r(x,y)\right)^2 * w(x,y)}. \tag{13}$$

Hence we can precompute $\Phi_l(\cdot,\cdot)$ and $\Phi_r(\cdot,\cdot)$ and use the shift operation to get $\Phi_l(\cdot,\cdot;u)$ and $\Phi_r(\cdot,\cdot;u)$. Fig. 5 illustrates the computational steps for constructing the SDS image defined by (2).

Procedure 1. $I^l, I^r \rightarrow \Xi$.

(1) Compute $\Phi_l(\cdot,\cdot)$:

$$\Phi_l(x,y) = \sqrt{\left(I^l(x,y)\right)^2 * w(x) * w(y)}.$$

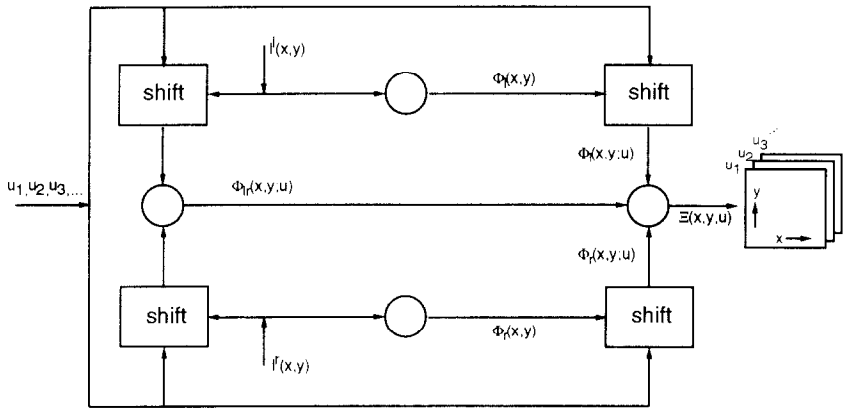

Fig. 5. Computing the SDS image defined using normalized correlation.

(2) Compute $\Phi_r(\cdot,\cdot)$:

$$\Phi_r(x,y) = \sqrt{\left(I^r(x,y)\right)^2 * w(x) * w(y)}.$$

(3) For each u do the following:

(3.1) Compute $\Phi_{lr}(\cdot,\cdot;u)$:

$$\Phi_{lr}(x,y;u) = \left(I^l(x+u/2,y)I^r(x-u/2,y)\right) * w(x) * w(y);$$

(3.2) Compute $\Xi(\cdot,\cdot,u)$:

$$\Xi(x,y,u) = \frac{\Phi_{lr}(x,y;u)}{\Phi_l(x+u/2,y)\Phi_r(x-u/2,y)}.$$

3. A surface-sensitive filter

3.1. Disparity surface in the SDS image

In the preceding section we have discussed the computation of the SDS image from stereo images. By computing the SDS, we reduce the problem of stereo matching to *feature detection*. The feature to be detected is the "bright" disparity surface in the three-dimensional SDS image. An SDS image can be regarded as a sequence of image slices $\Xi_y(x,u)$ indexed by y. As far as an image slice is concerned, the feature of interest is the "bright" disparity curve.

As an image feature, the disparity surface in the SDS image has the following geometric and photometric properties:

(1) *Explicit definition.* The disparity surface in $\Xi = \Xi(x,y,u)$ can be explicitly defined as a function $u = U(x,y)$. In the image slice $\Xi = \Xi(x,u)$ the disparity curve can be expressed as $u = U(x)$, and in the image slice $\Xi = \Xi(y,u)$, as $u = U(y)$.

Thus the disparity curve is more restrictive than conventional curves, e.g. highways in aerial images. For conventional curves the two image coordinates are usually related by *implicit* functions.

(2) *Weak continuity.* The disparity surface is continuous in the sense that in a neighborhood Ω_{xy} of (x, y) most points have similar disparities, i.e., $U \approx$ const. $\forall(\xi, \eta) \in \Omega_{xy}$. Due to the cohesive nature of objects in the physical world, $U(x, y)$ usually varies slowly with x and y. This property will enable us to distinguish the true bright points in the SDS image from ambiguous ones.

However, the disparity surface is in general non-flat. Although in practice the disparity surface gradient $\nabla U(x, y)$ is often close to zero, the flat surface assumption is misleading, especially when a large spatial neighborhood is concerned.

(3) *Equal brightness.* The SDS image intensity is close to the upper bound $\Xi_{\max}$ for a point on the disparity surface. In other words, disparity surface points are not only "bright" but also "equally bright" (ideally speaking). In contrast, in conventional curve detection the curve points usually have considerable intensity variations.

As indicated in the last section the SDS image is ambiguous; otherwise we would be able to detect the disparity surfaces simply by maximum-picking. To address this problem, we employ a filtering based approach. We want to design a filter which can effectively suppress the ambiguity and at the same time, enhance the disparity surface. The filter is supposed to take into account the above properties.

3.2. *Definition of the filter*

Let Ω_{xyu} be the receptive field associated with an SDS point (x, y, u). As shown in Fig. 6, Ω_{xyu} is a flat cube centered at (x, y, u). It can be expressed as

$$\Omega_{xyu} = \Omega_{xy} \otimes \Omega_u, \tag{14}$$

where $\otimes$ denotes "direct product". The size of Ω_{xyu} is $D_s \times D_s \times D_d$.

When applying the proposed filter ϕ to an SDS image, the response at (x, y, u) is computed from $\{\Xi(x, y, u)\colon (x, y, u) \in \Omega_{xyu}\}$. If Ω_{xyu} happens to capture a piece of the bright disparity surface, then $(\phi \circ \Xi)(x, y, u)$ is expected to be large; otherwise $(\phi \circ \Xi)(x, y, u)$ is expected to be small, even if (x, y, u) was bright before the filtering.

Consider the surface patch $u = U(x, y)$ defined on Ω_{xy}. If the patch is captured by the receptive field Ω_{xyu}, then it can be detected by using the following formula:

$$U(\xi, \eta) = \max_{\mu} \Xi(\xi, \eta, \mu), \quad (\xi, \eta, \mu) \in \Omega(x, y, u). \tag{15}$$

With the search space limited to the vicinity of u, spurious peaks can be avoided. This suggests that we can compute the response at (x, y, u) in two steps:

(1) detect a candidate surface patch defined on Ω_{xy} that is totally confined within the receptive field Ω_{xyu};

(2) accumulate neighborhood support by adding together the SDS image intensities of the points on the candidate surface patch.

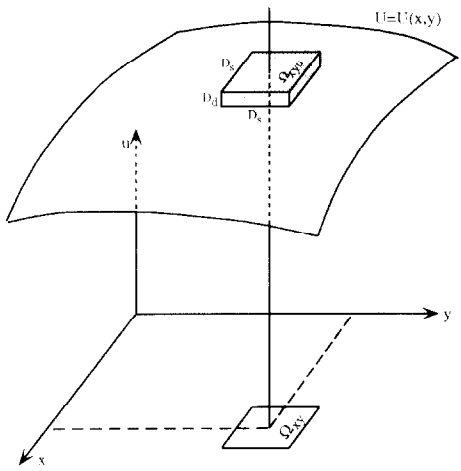

Fig. 6. The disparity surface and a "receptive field".

The filter is formally defined as follows. Given an SDS image $\Xi = \Xi(x, y, u)$, the response is

$$(\phi \circ \Xi)(x, y, u) = \sum_{\xi,\eta} g(\xi - x, \eta - y) \max_{\mu} \Xi(\xi, \eta, \mu), \quad (\xi, \eta, \mu) \in \Omega_{xyu}, \tag{16}$$

where $g(\cdot, \cdot)$ is a weighting function which is

(1) separable: $g(\xi, \eta) = g(\xi)g(\eta)$,

(2) symmetric: $g(\xi) = g(-\xi)$, and

(3) normalized.

In the case of disparity curve detection, ϕ is given by

$$(\phi \circ \Xi)(x, u) = \sum_{\xi} g(\xi - x) \max_{\mu} \Xi(\xi, \mu), \quad (\xi, \mu) \in \Omega_{xu}.$$

Fig. 7 illustrates the behavior of ϕ. The two gray rectangles are Ω_{xu} and $\Omega_{xu'}$ respectively. The support for (x, u) from ξ is $\Xi(Q)$, rather than $\Xi(\xi, u)$, where Q is found by searching in $\overline{cc'}$.

The proposed filter has the following properties:

- ϕ maps one SDS image Ξ to another SDS image $(\phi \circ \Xi)$. If (x, y, u) lies on the disparity surface, then Ω_{xyu} captures a surface patch on which the intensity is close to $\Xi_{\max}$. Hence $(\phi \circ \Xi)(x, y, u) \approx \Xi_{\max}$
- Lacking neighborhood support, spurious bright points in the SDS image becomes less bright after filtering.
- ϕ is a nonlinear filter; the response $(\phi \circ \Xi)(x, y, u)$ cannot be expressed as a linear combination of the intensity values in the receptive field. Obviously, the nonlinearity comes from the "max" operation.

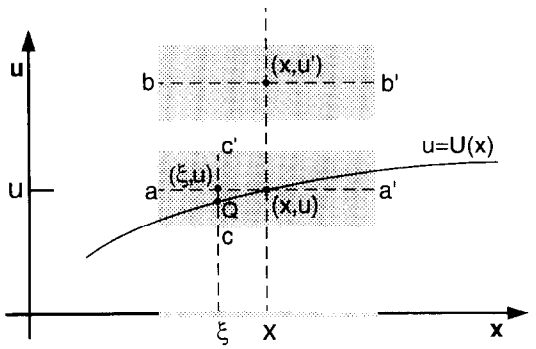

Fig. 7. The nonlinear filter ϕ.

3.3. Decomposition of the filter

The proposed filter is in fact a cascade of two well-known subfilters:

$$\phi = \phi_s \circ \phi_d,$$

where the first subfilter ϕ_d takes effect in the disparity domain and is a morphological filter; the second subfilter ϕ_s takes effect in the spatial domain and is a linear filter. Either of the subfilters is widely used in computer vision and image processing. Let focus on them one by one.

(1) *Disparity domain subfilter* ϕ_d. For each fixed (x, y) we have a 1D image $\Xi_{xy}(u) = \Xi(x, y, u)$. Applying ϕ_d amounts to computing the morphological dilation [30] of $\Xi_{xy}(u)$. Let e be the following structuring element:

$$e(u) = \begin{cases} 0, & \text{if } -D_d/2 \leqslant u \leqslant D_d/2, \\ 1, & \text{otherwise.} \end{cases}$$

Then the response to this subfilter is

$$\begin{aligned}(\phi_d \circ \Xi)(x, y, u) &= \Xi_{xy}(u) \oplus e(u) \\ &= \max_{\mu \in [u - D_d/2, u + D_d/2]} \Xi(x, y, \mu),\end{aligned} \tag{17}$$

where x and y are dummy variables.

(2) *Spatial domain subfilter* ϕ_s. Let $\mathcal{F}(x, y, u) = (\phi_d \circ \Xi)(x, y, u)$. For each fixed u we have a 2D image $\mathcal{F}_u(x, y) = \mathcal{F}(x, y, u)$. Applying ϕ_s amounts to convolving $\mathcal{F}_u(x, y)$ with $g(x, y)$:

$$\begin{aligned}(\phi_s \circ \mathcal{F})(x, y, u) &= \mathcal{F}_u(x, y) * g(x, y) \\ &= \sum_{(\xi, \eta) \in \Omega_{xy}} g(\xi - x, \eta - y) \mathcal{F}(\xi, \eta, u,)\end{aligned} \tag{18}$$

where u is a dummy variable.

From (17) and (18), we can easily verify that $\phi_s \circ \phi_d$ is indeed ϕ. Therefore, ϕ first computes morphological dilation in the disparity direction and then, convolution in the spatial directions. Note that ϕ_d precedes ϕ_s; the order cannot be reversed.

The decomposition provides an efficient way of applying ϕ. Obviously, calculating $\phi \circ \Xi$ directly from the definition (16) is computationally expensive. Thanks to the decomposition, we can first compute the morphological dilation in the disparity direction and then, the convolution along the spatial directions. By the way, the 2D spatial convolution with $g(x, y)$ can be computed by using 1D convolution operations, because g is assumed to be separable.

4. Multilevel surface enhancement

4.1. SDS image hyperpyramid

We have proposed a nonlinear filter for enhancing the disparity surface in the SDS image. A problem that needs to be addressed concerns the scale of the filter. On the one hand, the receptive field should preferably be big so as to capture a large surface patch so that sufficient image information can solicited for disambiguation. On the other hand, a big receptive field would complicate the use of the weak continuity assumption. In this section we will present a recursive filtering solution. The idea is to apply a sequence of successively bigger filters to the SDS image, which is equivalent to successively applying the same filter and shrinking the image, because the scale of the filter is related to that of the image. The result is a *hyperpyramid* representation of the SDS image.

Given an SDS image Ξ, the hyperpyramid $\{\Xi_m\}$ is defined as

$$\Xi_m(x, y, u) = (\phi \circ \Xi_{m-1})(2x, 2y, 2u), \quad \forall m > 1, \tag{19}$$
$$\Xi_1(x, y, u) = \Xi(x, y, u).$$

Conceptually, constructing Ξ_m from Ξ_{m-1} involves first applying ϕ. The result is then down-sampled by a factor of 2 in all the spatial and disparity directions. Note that the filter, as well as the down-sampling factor, is level-independent.

4.1.1. Nonlinearity of the hyperpyramid

Unlike the Gaussian pyramid, the SDS image hyperpyramid is nonlinear. Clearly Ξ_m, $m > 2$, cannot be computed from Ξ_1 in closed form; they are recursively defined. The hyperpyramid is for defining Ξ_m, rather than merely for speeding up computation.

Let us interpret what $\Xi_m(x, y, u)$ measures. We know that $\Xi(x, y, u)$ measures the confidence that (x, y, u) is on the disparity surface. In Ξ_2, a point (x, y, u) corresponds to

$$(x', y', u') = (x/2, y/2, u/2)$$

in Ξ, and $\Xi_2(x, y, u)$ measures the confidence that in Ξ, the "receptive field" of (x', y', u') of size

$$|\Omega^1_{x'y'u'}| = D_s \times D_s \times D_d$$

captures a patch of the disparity surface.

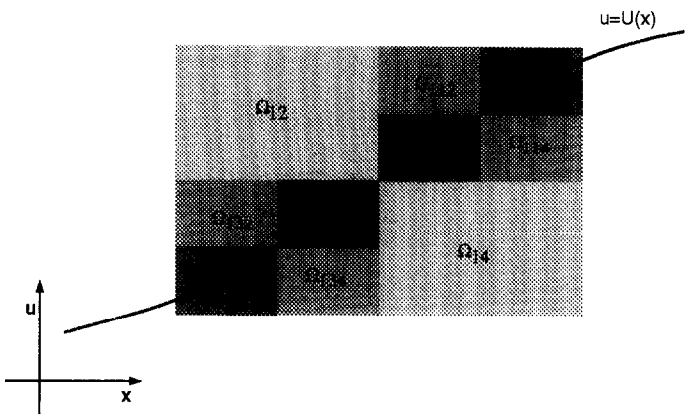

Fig. 8. Receptive fields in the xu-space.

Generically, a point (x, y, u) in Ξ_m, $m > 2$ corresponds to

$$(x', y', u') = (2^{-m+1}x, 2^{-m+1}y, 2^{-m+1}u)$$

in Ξ, and $\Xi_m(x, y, u)$ measures the confidence that in Ξ, the "receptive field" of (x', y', u') of size

$$|\Omega^{m-1}_{x'y'u'}| = 2^{m-2}D_s \times 2^{m-2}D_s \times 2^{m-2}D_d$$

captures a patch of the disparity surface.

Fig. 8 shows a disparity curve $u = U(x)$ and a number of receptive fields represented by rectangles. The big rectangle in the figure is denoted Ω_1 which is comprised of four subrectangles $\Omega_{11}, \Omega_{12}, \Omega_{13}$, and Ω_{14}, where Ω_{1i} is located at the ith quadrant of Ω_1. The subrectangle Ω_{11} again has four subrectangles $\Omega_{111}, \Omega_{112}, \Omega_{113}, \Omega_{114}$.

The receptive fields Ω_{111}, Ω_{113}, Ω_{131} and Ω_{133} all capture disparity curve segments. Thus the corresponding points in Ξ_2 are "bright". Since Ω_{11} captures Ω_{111} and Ω_{133}, the corresponding point in Ξ_3 looks "bright". By the same token, Ω_{13} projects to a "bright" point in Ξ_3. With information about the disparity curve being integrated, Ξ_2 is less ambiguous than Ξ_1 and Ξ_3 is less ambiguous than Ξ_2.

4.2. Computing the hyperpyramid

In the following we will present an algorithm for computing the SDS image hyperpyramid. The algorithm is simple and efficient; the basic operations are 1D convolution and local maximum-picking.

4.2.1. Receptive field size

In constructing the SDS hyperpyramid, we need to determine the size of the support of ϕ or the size of the receptive field $|\Omega_{xyu}| = D_s \times D_s \times D_u$. The receptive field has two parameters D_s and D_d. Their ratio D_s/D_d is a constant determined by experiments. We found that the ratio is around 5.5.

The receptive field is not allowed to contain only one pixel in the disparity direction. Otherwise, $\phi = \phi_s \circ \phi_d$ degenerates to ϕ_s. The *minimal receptive field* contains 2 pixels in the disparity direction.

With the use of the minimal receptive field, $\Xi_m(x, y, u)$ can be rewritten as follows:

$$\Xi_m(x, y, u) = \sum_{\xi,\eta} g(\xi - x, \eta - y) \max_{\mu=\{0,1\}} \Xi_{m-1}(2x, 2y, 2u + \mu), \quad (\xi, \eta) \in \Omega_{xy}, \tag{20}$$

where along the u-direction the indices of pixels start from zero.

The minimal receptive field principle leads to computational efficiency and gives good results. On the one hand, in computing Ξ_2 from $\Xi_1 = \Xi$ the weak continuity constraint is respected. On the other hand, recursive filtering enables us to eliminate the ambiguities remaining in Ξ_2.

4.2.2. Computation

We know that ϕ can be realized by cascading the two subfilters ϕ_d and ϕ_s. To derive Ξ_m from Ξ_{m-1}, we first apply the subfilter ϕ_d and down-sample in the u-direction at the same time. Then we apply the subfilter ϕ_s and down-sample in the spatial directions. Applying the subfilter ϕ_s amounts to convolution using the kernel $g(x, y)$. Since $g(x, y) = g(x)g(y)$ is separable, the 2D convolution can be computed efficiently using 1D convolution operations.

Fig. 9 shows the locations of the nodes of Ξ_m relative to those of Ξ_{m-1}. The nodes of Ξ_{m-1} are represented by "○", and the nodes of Ξ_m are represented by "●". Computing Ξ_m involves three steps:

$$\Xi_{m-1} = \Xi^{---} \rightarrow \Xi^{--+} \rightarrow \Xi^{-++} \rightarrow \Xi^{+++} = \Xi_m,$$

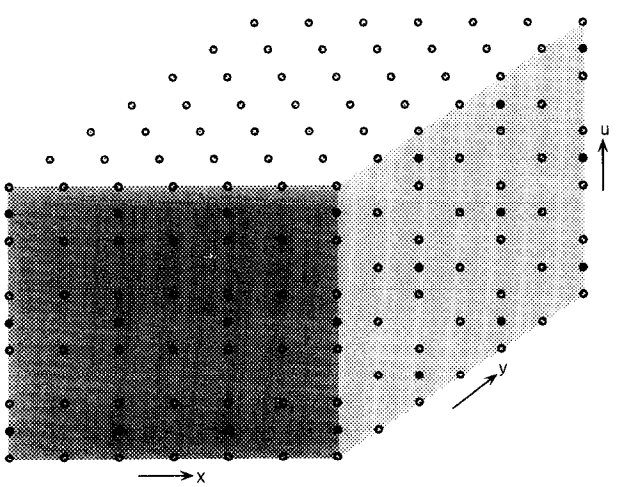

Fig. 9. Grid points of Ξ_{m-1} and Ξ_m.

Table 1
Sizes of the intermediate images

	$\Xi_{m-1} = \Xi^{---}$	Ξ^{--+}	Ξ^{-++}	$\Xi^{+++} = \Xi_m$
size	$A \times A \times B$	$A \times A \times \frac{B}{2}$	$A \times \frac{A}{2} \times \frac{B}{2}$	$\frac{A}{2} \times \frac{A}{2} \times \frac{B}{2}$

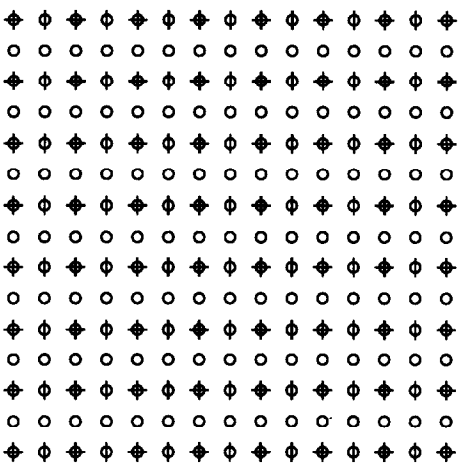

Fig. 10. Sampling the spatial domain.

where $-$ denotes pre-decimation and $+$ denotes post-decimation. Suppose Ξ_{m-1} is $A \times A \times B$, then the sizes of Ξ^{--+}, Ξ^{-++} and Ξ_m are as shown in Table 1. These intermediate images are explained as follows:

- $\Xi^{--+}(x, y, u)$ is simply the maximum of $\Xi_{m-1}(x, y, 2u)$ and $\Xi_{m-1}(x, y, 2u+1)$. The nodes of $\Xi^{--+}(x, y) = \Xi^{--+}(x, y, u)$, where u is a dummy variable, are shown as "o" in Fig. 10.
- Ξ^{-++} is the convolution of $\Xi^{--+}(x, y, u)$ and $g(y)$, computed at the locations marked by "|" in Fig. 10. It is worth emphasizing that, instead of first convolving and then down-sampling, we compute the convolution at predefined locations.
- $\Xi^{+++} = \Xi_m$ is the convolution of $\Xi^{-++}(x, y, u)$ and $g(x)$, computed at the locations marked by "–" in Fig. 10.

The procedure for constructing the SDS image hyperpyramid is as follows.

Procedure 2. $\Xi \rightarrow \{\Xi_m\}$.

(1) Initialize: $m = 1$; $\Xi_m = \Xi$.

(2) Increment: $m \leftarrow m + 1$.

(3) Compute Ξ_m from Ξ_{m-1}:

$$\Xi^{--+}(x, y, u) = \max_{\mu=\{0,1\}} \Xi_{m-1}(x, y, 2u + \mu),$$
$$\Xi^{-++}(x, y, u) = \big(\Xi^{--+}(x, y, u) * g(y)\big)(x, 2y, u),$$
$$\Xi_m(x, y, u) = \big(\Xi^{-++}(x, y, u) * g(x)\big)(x, 2y, u).$$

(4) If $m < M$ go to Step 2; otherwise stop.

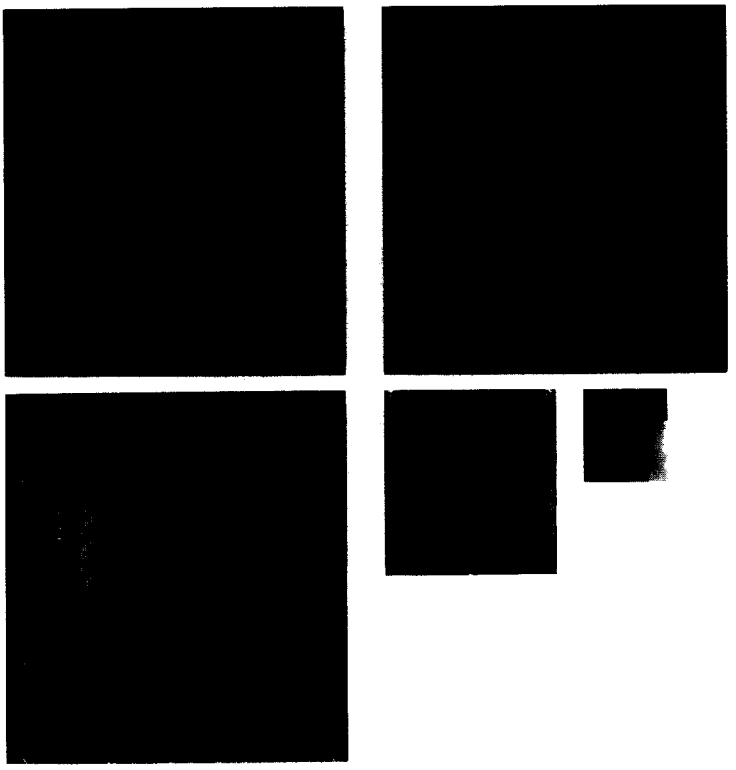

Fig. 11. "Parking meters" and results.

The above procedure is simple and computationally efficient. Computing an intensity $\Xi_m(x,y,u)$, $m > 1$, involves only three one-dimensional convolution operations. Moreover, $\Xi^{--+}(x,y,u)$ can be computed in parallel from Ξ_{m-1} with respect to (x,y,u). Similarly, $\Xi^{-++}(x,y,u), \forall(x,y,u)$ can be computed in parallel, and $\Xi^{+++}(x,y,u), \forall(x,y,u)$ can be computed in parallel.

4.2.3. Results

The effectiveness of the surface-enhancing filter and the multilevel scheme is demonstrated by experimental results. Typical results for outdoor images are shown in Figs. 11–13. The first rows in these figures show the input stereo pairs, which are "parking meters" (obtained by T. Kanade at CMU), "Denver" (obtained at the U.S. Army Topographic Engineering Center), and "Pentagon" respectively. The stereo images are positioned side by side to facilitate human stereopsis.

For each stereo pair an SDS image was computed using the normalized correlation measure. Then a three-level hyperpyramid was constructed. The disparity maps were computed at the three levels using maximum-picking. The results are shown in the second rows of Figs. 11–13 as intensity images, with disparity represented by brightness. All these results were obtained using the same parameters. As we can see, the ambiguities are reduced as we go from one level to the next, coarser level.

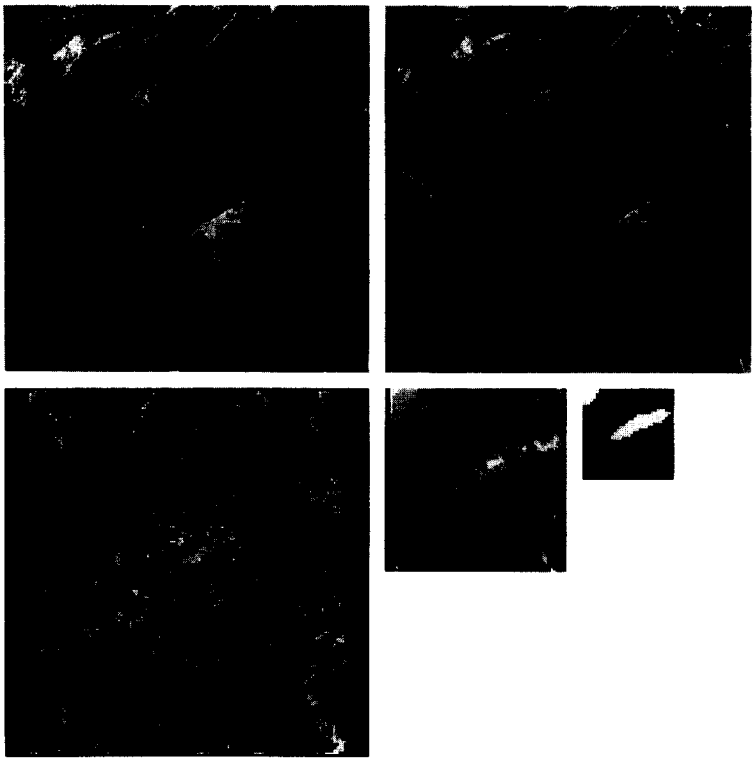

Fig. 12. "Denver" and results.

4.3. A degenerate case

Consider the degenerate case that in the u-direction the receptive field contains only one pixel. In this case, ϕ_d reduces to the identity operator and $\phi = \phi_s \circ \phi_d$ reduces to ϕ_s, and the SDS image hyperpyramid becomes a linear hyperpyramid.

Suppose that the SDS image is defined in terms of intensity difference. Then (1) detecting the disparities from Ξ_2 by $U(x, y) = \arg\max_u \Xi_2(x, y, u)$ is equivalent to the classical region matching approach to stereo correspondence, the region size being $D_s \times D_s$; and (2) detecting the disparities from Ξ_m, $m > 2$, is equivalent to region matching using a large, $2^{m-2}D_s \times 2^{m-2}D_s$ window. Note that implementation of the coarse-to-fine multiresolution method can be based on the degenerate SDS image hyperpyramid.

5. Multilevel surface detection

5.1. Coarse-to-fine approach

In the preceding section we discussed the computation of the SDS image hyperpyramid. In this section we address the problem of computing disparity surface from the hyperpyramid.

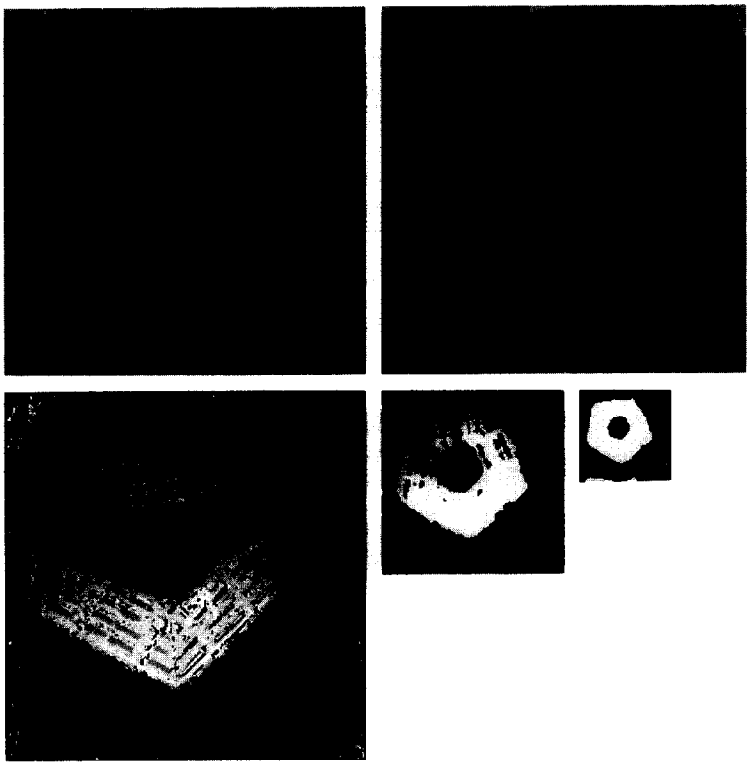

Fig. 13. "Pentagon" and results.

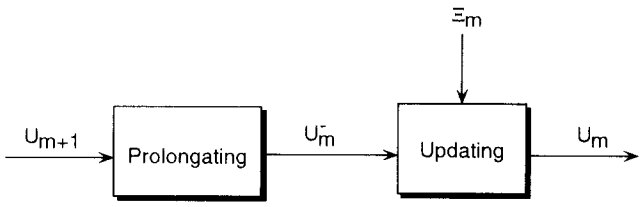

Fig. 14. Block diagram of inter-level recursion.

We address this problem using a coarse-to-fine approach. We first detect the disparity surface in the SDS image at the coarsest level, then in the SDS image at the next, finer level. This process is repeated until the finest level is reached. In this way, a sequence of of surfaces $U_M, \ldots, U_2, U_1$ are generated, in this order. When detecting U_m in Ξ_m, $m < M$, the information provided by U_{m+1} is utilized.

Computing U_m, $m < M$ involves two steps: one is *prolongation* and the other is *updating*, as shown in Fig. 14. The prolongation operation produces U_m^- which is an expanded version of U_{m+1}. The updating step can proceed in many ways. For example, one may employ a Bayesian approach [31] and modify U_m^- iteratively to minimize

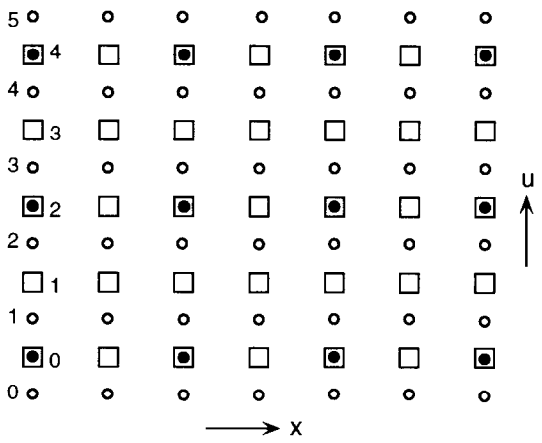

Fig. 15. Spaces of U_m^- and Ξ_m.

some error functional. But this approach is computationally inefficient. As it is to be shown, the hyperpyramid allows us to detect the stereo disparities in a recursive but non-iterative way.

5.1.1. Prolongation

In proceeding from one level $m+1$ from the next level m, the first step is to prolongate U_{m+1} to derive a rough estimate U_m^- of the disparity surface in Ξ_m. A simple way to derive U_m^- is to use bilinear interpolation.

Fig. 15 illustrates the grid points of Ξ_{m+1}, represented by •, and those of Ξ_m, represented by ○. If x and y are even, then $U_m^-(x,y) = 2U_{m+1}(x/2, y/2)$. Generically, $U_m^-(x,y)$ can be expressed as

$$U_m^-(x,y) = \tfrac{1}{2}\big(U_{m+1}(\lceil x/2\rceil, \lceil y/2\rceil) + U_{m+1}(\lceil x/2\rceil, \lfloor y/2\rfloor) + U_{m+1}(\lfloor x/2\rfloor, \lceil y/2\rceil) + U_{m+1}(\lfloor x/2\rfloor, \lfloor y/2\rfloor)\big), \tag{21}$$

where $\lceil\ \rceil$ and $\lfloor\ \rfloor$ denote "ceil" and "floor" respectively. The right-hand side of (21) may yield a non-integer when both x and y are not even. We let $U_m^-(x,y)$ be the integer closest to the right-hand side value. In other words, *integer interpolation* is performed.

The possible locations of $U_m^-(x,y)$ are represented by □ in Fig. 15. The space of $U_m^-(x,y)$ is a superset of that of $U_{m+1}(x,y)$, which is represented by •. Note that $(x, y, U_m^-(x,y))$ is not a grid point of Ξ_m represented by ○.

5.1.2. Updating

Once U_m^- has been obtained via interpolation, the updating step that computes U_m can simply be local maximum-picking. Fig. 15 shows the locations of U_m^- (represented by □) relative to the voxels of Ξ_m (represented by ○). To locate $U_m(x,y)$ we may search the following $2k$ squares:

$$\mathcal{U} = \{U_m^-(x,y) - (k-1), \ldots, U_m^-(x,y) + k - 1, U_m^-(x,y) + k\}, \tag{22}$$

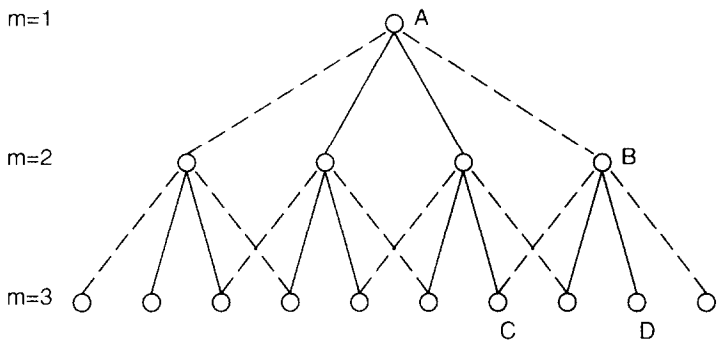

Fig. 16. Tree search in the coarse-to-fine sweep.

where the disparity indices of U_m^- and Ξ_m start from zero. The local search can be expressed as:

$$U_m(x,y) = \arg\max_{\mu\in\mathcal{U}} \Xi_m(x,y,\mu) \tag{23}$$

In order to avoid ambiguous peaks, k in (22) has to be small. Possible choices include $k = 1$ and $k = 2$.

Obviously $k = 1$ or $\mathcal{U} = \{U_m^-(x,y), U_m^-(x,y) + 1\}$ is infeasible. If $k = 1$, then the accuracy at the finest level is determined by that at the coarsest level. In Fig. 16, A represents the disparity of a point (x,y) in Ξ_3. Note that $k = 1$ implies the search of the binary tree search rooted at A. If D the true disparity of the point $(4x,4y)$ in Ξ_1, then it can never be reached by the binary search. In fact, all the nodes to the right of node C cannot be reached.

We let $k = 2$, which implies a 4-ary tree search. In Fig. 16, the four children of A are compared and the one which is the brightest is taken to be the disparity of the point $(2x,2y)$ in Ξ_2. Suppose that B is the winner. As we proceed from level 2 to level 1, the four children of B are visited and the winner locates the disparity of $(4x,4y)$ in Ξ_1.

In other words, U_m^- specifies a local search space

$$\mathcal{U}_m = \{U_m^-(x,y) - 1, U_m^-(x,y), U_m^-(x,y) + 1, U_m^-(x,y) + 2\},$$

and the updating step performs maximum-picking:

$$U_m(x,y) = \arg\max_{\mu\in\mathcal{U}_m} \Xi_m(x,y,\mu). \tag{24}$$

Now it become clear why we used integer interpolation in computing U_m^- from U_{m+1} by (21). Indeed, $U_m^-(x,y)$ is used to identify the four points (or the four □s in Fig. 15) among which the maximum is to be picked.

5.2. Computation

The disparity surface is computed from the SDS image pyramid in the following steps.

Procedure 3. $\{\Xi_m\} \rightarrow U$.

(1) Initialize: Let $m = M$ and compute

$$U_m(x, y) = \arg\max_u \Xi_m(x, y, u).$$

(2) Recurse: Decrement $m \leftarrow m - 1$.
  (2.1) Prolongation: Compute U_m^- from U_{m+1} using (21).
  (2.2) Updating: Compute U_m using (24).

(3) Terminate: If $m > 1$ go to Step 2; otherwise let $U = U_1$ and stop.

The above procedure is simple and efficient. The basic operation is maximum-picking from a 4-element set, and control structure is coarse-to-fine level-recursive, with no intra-level iterations. We will demonstrate that such a simple algorithm produces good results.

5.3. Results

In this subsection we will present the results of computing disparities from SDS image hyperpyramids (for details, see [35]). We will demonstrate that the simple method described by Procedure *CtF2* gives good results. All the disparity values are accurate to 1 pixel.

Figs. 17–19 show the disparity maps for "parking meters", "Denver", and "Pentagon", respectively. By comparing Fig. 17 with the second row of Fig. 11 we can see that local search is effective in resolving ambiguities.

Fig. 17 shows that the algorithm has captured the three parking meters in front of the shrubberies, the side view of the signboard which is between the second and the third parking meters, and the large depth gap between the front and the building.

Fig. 18 shows that the general relief of the terrain is recovered. Fig. 19 shows that the complex roof structure is well recovered. Moreover, the road and overpass on the right-hand side of the scene is well modeled. The results for "Denver" and "Pentagon" compare favorably to those obtained by Hsieh et al. [12] using the multiresolution method.

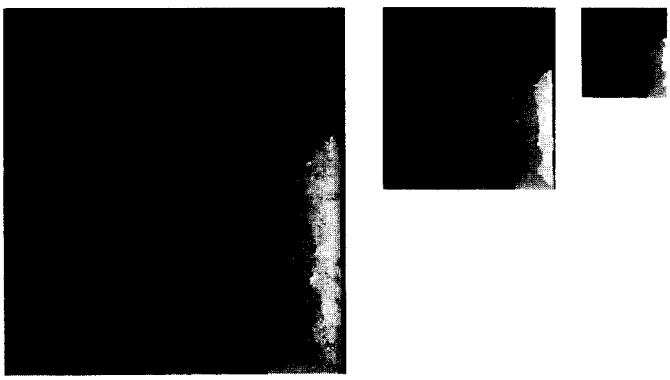

Fig. 17. Results for "parking meters".

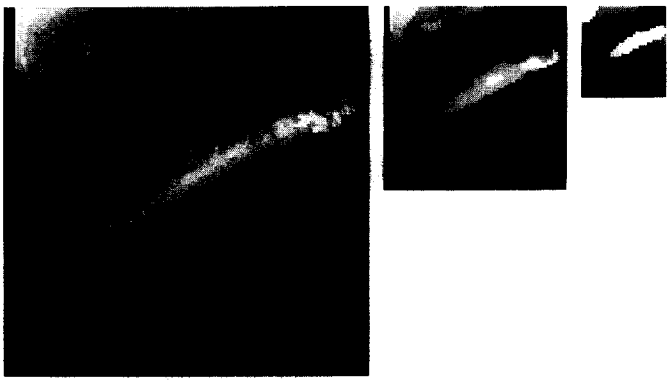

Fig. 18. Results for "Denver".

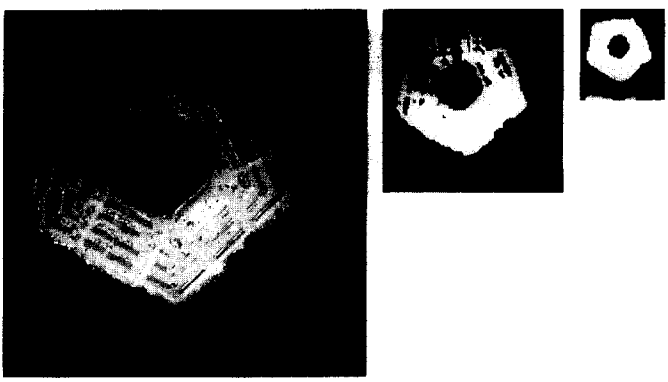

Fig. 19. Results for "Pentagon".

The results shown in Figs. 17–19 are represented with respect to the cyclopean view. In other words, they are the cyclopean disparity maps. We can also obtain the left disparity maps and the right disparity maps (cf. (4) and (5)). The first row Fig. 20 shows the left disparity map U^l and the (negative) right disparity map $-U^r$. Fig. 20 (lower-right) shows the regions visible from the left camera only. The left-visible regions were detected by checking the consistency between U^l and U^r. (For binocularly visible points, $U^l(x,y) - U^r(x - U^l(x,y),y) \equiv 0$. Fig. 20 (lower-left) shows the left disparity map with the left-visible regions being shown in black.

6. Summary and concluding remarks

The notation of spatio-disparity space, or xyu-space, is very useful for stereo vision. It is related to the commonly used notion of matching space, or the $x^l x^r y$-space, where each point (x^l, x^r, y) represents a match between (x^l, y) in the left image and (x^r, y) the right image. The SDS representation is desirable because the output of the stereo

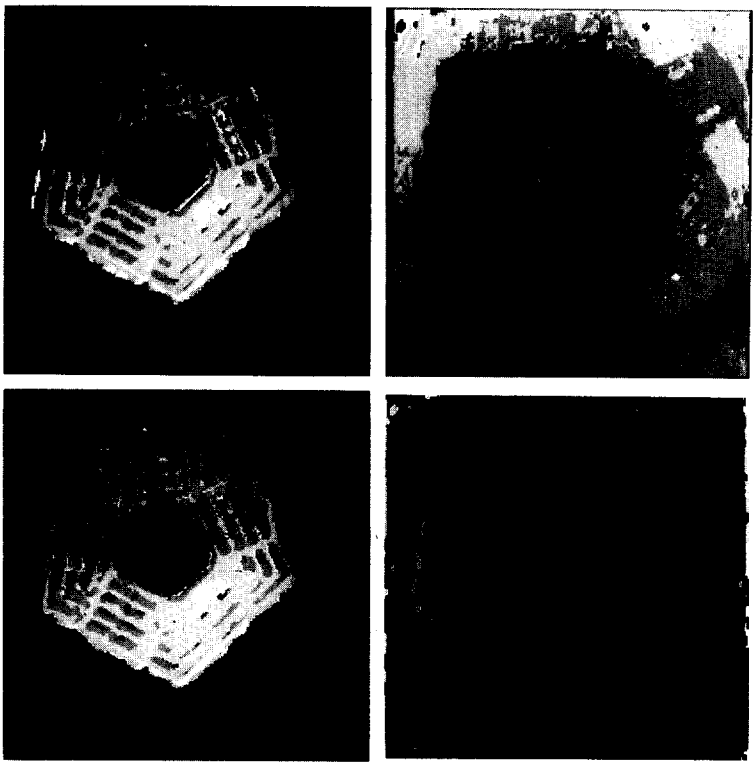

Fig. 20. Results of occlusion detection for "Pentagon".

algorithm gives the disparity as a function of the spatial position. The SDS image is a useful representation of stereo information in the left and right images.

By computing the SDS image, we reduce the problem of stereo matching to feature detection, where the feature is the disparity surface in the SDS image $\Xi = \Xi(x, y, u)$, or the disparity curve in $\Xi = \Xi(x, y)$. In fact, generic image matching can be formulated as feature detection. For example, the computation of optical flow $\boldsymbol{U} = \boldsymbol{U}(x, y)$ can be formulated as the detection of $\boldsymbol{u} = \boldsymbol{U}(x, y)$ in the *spatio-velocity space image* $\Xi = \Xi(x, y, \boldsymbol{u})$ where $(x, y, \boldsymbol{u})$ denotes the match between $(x, y) + \boldsymbol{u}/2$ in the first frame of image and $(x, y) - \boldsymbol{u}/2$ in the second frame.

We have proposed a filter ϕ for detecting the disparity surface in the SDS image. By applying ϕ, the disparity surface is enhanced and the noise is suppressed. The design of ϕ has exploited the photometric and geometric attributes of the disparity surface. Unlike many common feature detectors, ϕ is nonlinear by nature. The nonlinearity provides a neat way for accommodating the disparity surface undulations or nonlinearities. Being a novel filter *per se*, ϕ can decomposed into two subfilters, one of which being a morphological filter and the other being a linear filter. This decomposition leads to simple and efficient computation of disparity surface enhancement.

The multilevel paradigm for stereo matching can be described as a fine-to-coarse process followed by a coarse-to-fine process. In the fine-to-coarse sweep stereo information

is accumulated in an attempt to remove ambiguities. The SDS hyperpyramid provides a sequence of successively surface-enhanced versions of the SDS image. In the coarse-to-fine sweep, the disparity surface is detected at each level and is refined at the next, finer level. This fine-to-coarse-to-fine process is computationally efficient, involving no intra-level iterations.

To conclude, the multilevel method proposed here provides a novel hierarchical paradigm for computing stereo disparities. In addition to applications in cartography and robot vision, the multilevel algorithm may shed light on the understanding of the underlying biological mechanism of human stereo vision.

Acknowledgment

We are grateful to David Mumford and Jim Clark for their insights. We thank Roger Brockett, Petros Maragos, Song De Ma, and Jie Lu for invaluable discussions. We also thank anonymous reviewers for constructive comments which greatly enhanced the presentation of this paper.

This work was supported in part by the Brown/Harvard/MIT Center for Intelligent Control Systems with U.S. Army Research Office grant number DAAL03-86-K-0171.

References

[1] P. Anandan, A computational framework and an algorithm for the measurement of visual motion, *Int. J. Comput. Vision* **2** (1989) 283–310.

[2] H.H. Baker and T.O. Binford, Depth from edge- and intensity-based stereo, in: *Proceedings IJCAI-81*, Vancouver, BC (1981) 631–636.

[3] S.T. Barnard, Stochastic stereo matching over scale, *Int. J. Comput. Vision* **2** (1989) 17–32.

[4] S.T. Barnard and M.A. Fischler, Stereo vision, in: S.C. Shapiro, ed., *Encyclopedia of Artificial Intelligence* (1987) 1083–1090.

[5] P.N. Belhumeur and D. Mumford, A Bayesian treatment of the stereo correspondence problem using half-occluded regions, in: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition* (1992).

[6] P.J. Burt, Fast filter transformations for image processing, *Comput. Vision Graph. Image Process.* **16** (1981) 20–51.

[7] S.D. Cochran and G. Medioni, 3-D surface description from binocular stereo, *IEEE Trans. Pattern Anal. Mach. Intell.* **14** (1992) 981–994.

[8] W. Forstner and A. Pertl, *Photogrammetric Standard Methods and Digital Image Matching Techniques for High Precision Surface Measurements* (Elsevier Science Publishers B.V., Amsterdam, 1986).

[9] W.E.L. Grimson, *From Images to Surfaces* (MIT Press, Cambridge, MA, 1981).

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

[11] W. Hoff and N. Ahuja, Surface from stereo: integrating feature matching, disparity estimation, and contour detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **11** (2) (1989) 121–136.

[12] Y.C. Hsieh, D.M. McKeown Jr and F.P. Perlant, Performance evaluation of scene registration and stereo matching for cartographic feature extraction, *IEEE Trans. Pattern Anal. Mach. Intell.* **14** (2) (1992) 214–238.

[13] D.G. Jones and J. Malik, Determining three-dimensional shape from orientation and spatial frequency disparities, in: *Proceedings Second European Conference on Computer Vision*, Genova, Italy (1992).

[14] B. Julesz, *Foundations of Cyclopean Perception* (The University of Chicago Press, Chicago, IL, 1971).

[15] T. Kanade and M. Okutomi, A stereo matching algorithm with an adaptive window: theory and experiment, in: *Proceedings International Conference on Robotics and Automation*, Sacramento, CA (1991) 1088–1095.
[16] T. Kanade, M. Okutomi and T. Nakahara, A multiple-baseline stereo method, in: *Proceedings DARPA Image Understanding Workshop* (1992) 409–426.
[17] B.D. Lucas, Generalized image matching by the method of differences, Ph.D dissertation, Carnegie-Mellon University, Pittsburgh, PA (1984).
[18] D. Marr, *Vision* (Freeman, San Francisco, CA, 1982).
[19] D. Marr and T. Poggio, A theory for human stereo vision, *Proc. Roy. Soc. Lond. B* **204** (1979) 301–328.
[20] L. Matthies, Stereo vision for planetary rovers, *Int. J. Comput. Vision* **8** (1) (1992) 71–91.
[21] H.P. Moravec, Obstacle avoidance and navigation in the real world by a seeing robot rover, Ph.D. dissertation, Stanford University, Stanford, CA (1980).
[22] Y. Ohta and T. Kanade, Stereo by intra- and inter-scanline search using dynamic programming, *IEEE Trans. Pattern Anal. Mach. Intell.* **7** (1985) 139–154.
[23] M. Okutomi and T. Kanade, A locally adaptive window for signal matching, *Int. J. Comput. Vision* **7** (2) (1992) 143–162.
[24] M. Okutomi and T. Kanade, A multi-baseline stereo, *IEEE Trans. Pattern Anal. Mach. Intell.* **15** (4) (1993) 353–363.
[25] T. Poggio, V. Torre and C. Koch, Computational vision and regularization theory, *Nature* **317** (1985) 314–319.
[26] S.B. Pollard, J.E.W. Mayhew and J.P. Frisby, PMF: a stereo correspondence algorithm using a disparity gradient limit, *Perception* **14** (1985) 449–470.
[27] L.H. Quam, Hierarchical warp stereo, in: *Proceedings DARPA Image Understanding Workshop* (1984) 149–156.
[28] A. Rosenfeld, ed., *Multiresolution Image Processing and Analysis* (Springer-Verlag, New York, 1984).
[29] A. Rosenfeld, ed., Pyramid algorithms for global structures in images, *Inform. Sci.* **50** (1) (1990) 23–34.
[30] J. Serra, *Image Analysis and Mathematical Morphology* (Academic Press, London, 1982).
[31] J. Weng, N. Ahuja and T.S. Huang, Matching two perspective views, *IEEE Trans. Pattern Anal. Mach. Intell.* **14** (8) (1992) 806–825.
[32] C. Wheatstone, On some remarkable, and hitherto unobserved, phenomena of binocular vision, *Philos. Trans. Roy. Soc.* **128** (1836) 371–394.
[33] A. Witkin, D. Terzopoulos and M. Kass, Signal matching through scale space, *Int. J. Comput. Vision* **1** (1987) 133–144.
[34] Y. Yang, A.L. Yuille and J. Lu, Local, global, and multilevel stereo matching, in: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition* (1993) 274–279.
[35] Y. Yang, Multilevel computation of stereo correspondence, Ph.D. Thesis, Harvard University, Cambridge, MA (1994).