

# Stereo Matching as a Nearest-Neighbor Problem

Carlo Tomasi, *Member, IEEE*, and Roberto Manduchi

**Abstract**—We propose a representation of images, called *intrinsic curves*, that transforms stereo matching from a search problem into a nearest-neighbor problem. Intrinsic curves are the paths that a set of local image descriptors trace as an image scanline is traversed from left to right. Intrinsic curves are ideally invariant with respect to disparity. Stereo correspondence then becomes a trivial lookup problem in the ideal case. We also show how to use intrinsic curves to match real images in the presence of noise, brightness bias, contrast fluctuations, moderate geometric distortion, image ambiguity, and occlusions. In this case, matching becomes a nearest-neighbor problem, even for very large disparity values.

**Index Terms**—Stereo vision, stereo matching, correspondence problem, disparity, ambiguity, occlusions, search, nearest-neighbor search, dynamic programming.

## 1 INTRODUCTION

STEREO correspondence is traditionally associated with search: For every point in the left image, the right image is searched for a similar point. In this paper, we show that search is not inherent in the correspondence problem. Instead, matching becomes a nearest-neighbor problem. The way out of search is associative memory, and essentially inverts the way images are represented. Rather than storing image intensities by their position in the image, the usual array  $I(x)$ , we can store image positions by their appearance: in a sense,  $x(I)$ . Then, image points that look similar are stored in the same place. If both images are stored in the same memory, correspondences are trivially established, because corresponding points share the same memory locations. Occlusions are also easily found as points that live alone in some location. There are two problems with this scheme. Different image points can look the same, so memory locations can be crowded; and corresponding points in the two images can look different, so that points that ought to go in the same memory location do not.

We address the first problem by analyzing possible changes between corresponding scanlines, so we know where to look next if a memory location is missing a point. The second problem is addressed by a twofold strategy. First, it is reduced by encoding image appearance with richer descriptors than image intensity: Each image location is described by a vector of parameters. Second, the resolution of the remaining ambiguity is made easier by preserving contiguity information with the descriptors. Consider traversing a scanline in one of the two images. The vector of descriptors traces a curve in some space, and points that are nearby in the scanline are also nearby in the representation. Contiguity then helps selecting among similar match candidates: When two points look similar, we look around them and match entire curve segments at once, rather than isolated points.

To illustrate our approach, here is one simple version of what we call an *intrinsic curve* for, say, the left scanline in row 68 of the image in Fig. 1. A low-pass filtered version of the image intensity

• The authors are with the Computer Science Department, Stanford University, Stanford, CA 94305, USA. E-mail: {tomasi,manduchi}@cs.stanford.edu.

Manuscript received 9 Aug. 1996; revised 13 Jan. 1998. Recommended for acceptance by G. Medioni.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 106319.



Fig. 1. Test image "Trees" from SRI, frame 1.

$I(x)$  and its derivative  $I'(x)$  are computed everywhere (solid lines in Fig. 2b and Fig. 2c) and are plotted against each other (solid lines in Fig. 3a and Fig. 3b). When plotting  $I'$  versus  $I$ , we lose track of space, that is, of the coordinate  $x$  which merely parameterizes the curve  $I'(I)$ . This parameter is stored for later use, but it plays no role in the shape of the curve. If  $I(x)$  is replaced by a shifted replica  $r(x) = I(x + d)$ , the curve of Fig. 3b remains the same. The dashed curves in Fig. 2 and Fig. 3 show the construction of the intrinsic curve for the scanline  $r(x)$  taken from a different viewing position.

More general geometric transformations between  $l$  and  $r$ , and photometric distortions on the range of the two signals, can deform an intrinsic curve. In the next section, we present a theory of intrinsic curves, and study the effect of geometric and photometric distortions. Because of these distortions, matches are nearest neighbors in the space of descriptors. Our proposed metric in this space, an algorithm for the nearest-neighbor computation, and a dynamic programming technique for global optimization are described in Section 3 and tested on real stereo pairs.

## 2 INTRINSIC CURVES: THEORY

An efficient procedure for matching two signals is to compare vectorial descriptions of their local intensity variations [3], [4], [5]. In this section, we define intrinsic curves more generally. We also identify the geometric mappings

$$r(x) = I(\alpha(x)) \quad (1)$$

between the two images that are *compatible* with any particular way to build intrinsic curves, in the sense that they leave the curves unaltered. Finally, we investigate geometrical and topological properties of intrinsic curves.

**DEFINITION.** An intrinsic curve: Suppose that the  $N$  operators  $P_1, \dots, P_N$  are applied to the signal  $I(x)$  to produce the new signals

$$p_n(x) = [P_n I](x)$$

for  $n = 1, \dots, N$ . The vector

$$\mathbf{p}(x) = (p_1(x), \dots, p_N(x)) \quad (2)$$

describes a curve  $C$  in  $R^N$  parameterized by the real variable  $x$ :

$$C = \{\mathbf{p}(x), x \in R\}. \quad (3)$$

$C$  is called the intrinsic curve generated by  $I(x)$  through the operators  $P_1, \dots, P_N$ .

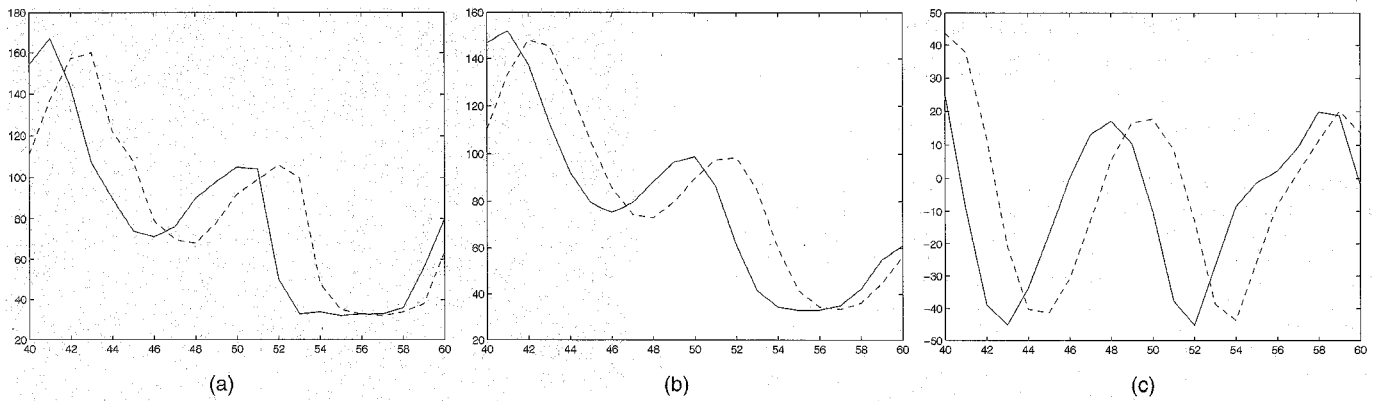


Fig. 2. Scanline 68 from the image of Fig. 1, pixels 40–60 (solid line: frame 1; dashed line: frame 2). (a) Intensity. (b) Its low-pass filtered version. (c) Its derivative.

**2.1 Compatible Mappings**

Any reparameterization  $x \rightarrow \alpha(x)$  of  $C$  leaves  $C$  unchanged. However, reparameterizing  $l(x)$  to  $l(\alpha(x))$  can, in general, modify  $C$ . For instance, if

$$p_1(x) = l(x) \quad \text{and} \quad p_2(x) = l'(x),$$

where the prime denotes differentiation, the new components of  $C$  after the change  $x \rightarrow \alpha(x)$  become

$$\tilde{p}_1(x) = l(\alpha(x)) \quad \text{and} \quad \tilde{p}_2(x) = \alpha'(x)l'(\alpha(x))$$

so that  $\tilde{p}(x)$  traces a new curve that is modulated by  $\alpha'(x)$  in its second component.

**DEFINITION.** A compatible mapping: A mapping  $x \rightarrow \alpha(x)$  is said to be compatible with the operators  $P_1, \dots, P_N$  if, for every signal  $l(x)$ , the intrinsic curve generated by  $l(x)$  is equal to the intrinsic curve generated by  $l(\alpha(x))$ .

The set of compatible mappings depends on the choice of the operators  $P_1, \dots, P_N$ . For instance, it is easy to verify that

- the constant displacements  $\alpha(x) = x + d$  are compatible with shift-invariant operators;
- the affine mappings of the form  $\alpha(x) = ax + d$  are compatible with the operators

$$p_n(x) = [P_n l](x) = \frac{\left(\frac{d^n}{dx^n} l(x)\right)^{(n+1)/n}}{\frac{d^{n+1}}{dx^{n+1}} l(x)}$$

defined wherever  $\frac{d^{n+1}}{dx^{n+1}} l(x) \neq 0$ ;

- if the mapping  $x \rightarrow \alpha(x)$  is regarded as an operator  $A$  applied to  $l(x)$ , that is, if

$$[Al](x) = l(\alpha(x)),$$

then  $\alpha(x)$  is compatible with operators  $P_1, \dots, P_N$  if there is an operator  $D$  such that for every  $n$ ,

$$P_n A = D P_n$$

where  $D$  is a diffeomorphism independent of  $n$ . Both previous examples are special cases of this general rule.

Thus, intrinsic curves can be regarded as invariants with respect to the set of compatible mappings, and provide a richer and more general description than “classical” invariants such as function moments [6], [7], [8], [9], affine mappings [10], [11], [12], and shift-invariant filters [3], [4], [13], [14], [15]. The fact that, in gen-

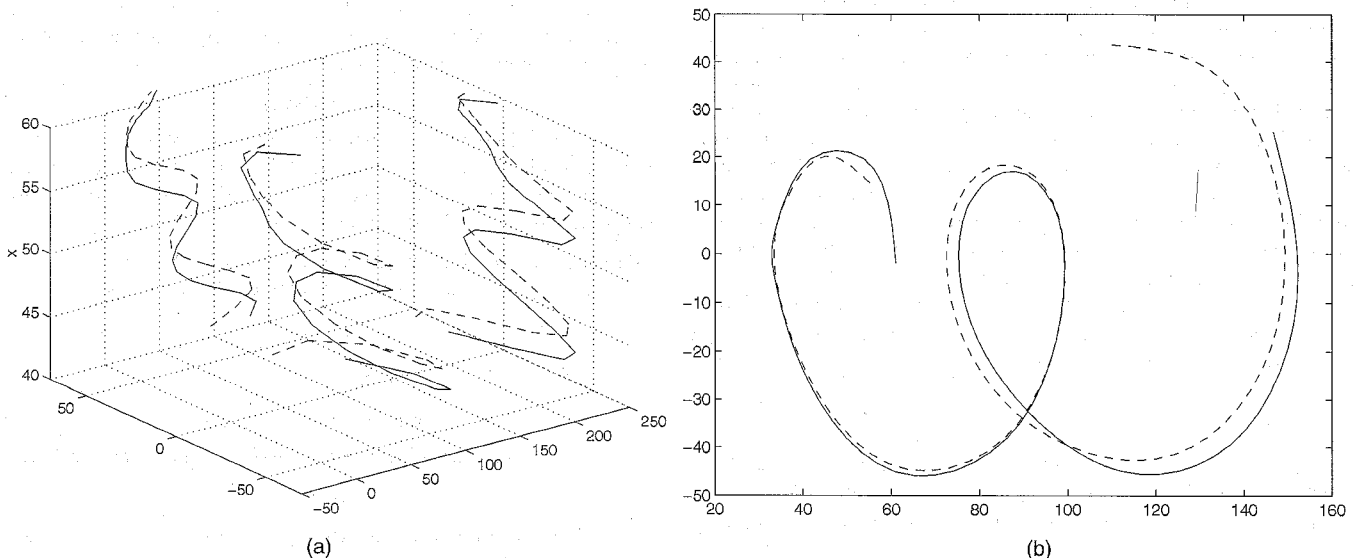


Fig. 3. (a) Intrinsic curve formation: The signals of Fig. 2b and Fig. 2c are plotted against each other, forming a 3D curve whose projection on the plane  $x = 0$  is the intrinsic curve (b).

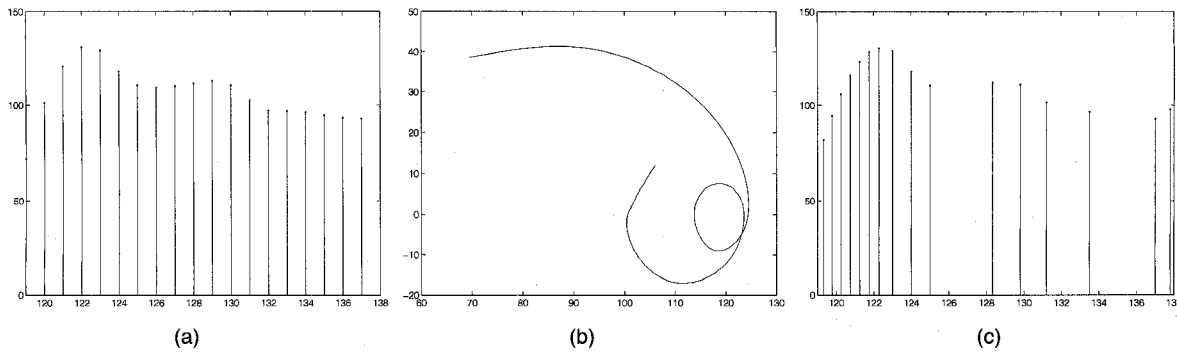


Fig. 4. A signal sampled on a uniform grid (a) and on a nonuniform grid (c) induced by the uniform arc length sampling of the intrinsic curve (b).

eral, affine mappings are not compatible with shift-invariant operators is, therefore, important and was also pointed out in [3]. In the remainder of this section, we assume that the mapping  $\alpha(x)$  is a diffeomorphism. In addition, we assume throughout this paper that both the input signals  $l(x)$ ,  $r(x)$  and the operators  $P_n$  are continuous.

If the transformation between the left and right images were just a mapping  $\alpha(x)$  compatible with the operators  $P_1, \dots, P_N$ , stereo matching would be nearly trivial, because the two intrinsic curves would coincide, and their parameterizations by  $x$  would yield the correspondences.

In reality, the left and right images in a stereo pair are related in a more complex way. First, the actual mapping  $\alpha(x)$  is at best only approximately compatible with any given set of intrinsic curve operators. Second, photometric distortions  $\phi(l(x))$  combine with geometric distortions  $l(\alpha(x))$ . Third, noise corrupts both  $l(x)$  and  $r(x)$ . Fourth, occlusions and specularities complicate matching.

In the next section, we analyze the properties of special intrinsic curves on the plane. Then, in Section 2.3, we analyze the discrepancies between corresponding intrinsic curves from real stereo pairs.

## 2.2 Properties of Intrinsic Curves

In this section, we concentrate on planar ( $N = 2$ ) intrinsic curves with

$$p_1(x) = [P_1]l(x) = l(x) \quad \text{and} \quad p_2(x) = [P_2]l(x) = l'(x), \quad (4)$$

reminiscent of the *phase space* diagrams of systems theory [16]. The following properties of intrinsic curves in the phase space are proven in [24].

The intrinsic curve generated by  $l(x)$  is connected iff the first derivative of  $l(x)$  is continuous. The curve is differentiable  $n$  times iff  $l(x)$  is differentiable  $n + 1$  times.

Intrinsic curves are traversed left to right in the upper half-plane and right to left in the lower as one traverses the original signal from left to right. Also, when crossing the axis of the abscissas, the curve tangent forms an angle of  $\pm \pi/2$  with it. If  $l(x)$  is zero mean, its intrinsic curve loops around the origin.

Imagine traversing the image scanline  $l(x)$  from left to right at constant speed in  $x$ . The corresponding point  $p$  in phase space then moves along the intrinsic curve  $(l(x), l'(x))$  at a velocity that depends on the local behavior of  $l(x)$ . The relationship between these two traversal velocities is shown in Fig. 4. If, in turn, we sample the curve with constant arc length period through spline interpolation, we induce a nonuniform sampling of the signal, as shown in Fig. 4c. Samples are denser where the signal busyness is higher, less dense where the signal is flat. This is a very desirable property for stereo matching, because matches are less robust where the image is "flat."

The trajectory  $p(x)$  may go through a point  $p_0$  more than once (self-intersect), when the local image description is the same at

more than one point  $x$ . Increasing the dimensionality of the local description provides a richer representation (as long as the operators are independent), as noticed also in [3], [4]. This becomes obvious with intrinsic curves. For example, with two operators the intrinsic curves lie in a plane, and self-intersections are to be expected. With three operators, the curves live in a 3D space, where intersections are less likely.

In more than two dimensions, we may consider intrinsic curves  $C$  in  $R^N$  of the form  $p(x) = (l(x), l'(x), l''(x), \dots, l^{(N)}(x))$ . The properties described above apply to the projection of  $C$  onto each plane  $(l^{(n)}, l^{(n+1)})$ . These definitions can be extended to negative values of  $n$  as well by defining

$$l^{(n)}(x) = \int_{-\infty}^x \int_{-\infty}^{x-n} \dots \int_{-\infty}^{x_2} l(x_1) dx_1 \dots dx_{-n} \quad \text{for } n < 0.$$

## 2.3 Deviations From the Ideal Case

Intrinsic curves of corresponding scanlines  $l(x)$  and  $r(x)$  related by a compatible mapping

$$r(x) = l(\alpha(x)) \quad (5)$$

are ideally identical. In reality, however, they can differ for the reasons explained below.

### 2.3.1 Occlusions

When parts of  $l(x)$  or  $r(x)$  are occluded,  $\alpha(x)$  is undefined. Occlusions are studied in [18], [19], [4], [20], [21], [22]. With intrinsic curves, an occlusion produces an arc of one curve that is not matched in the other. Just before and just after the unmatched arc, the curves are expected to coincide. See Fig. 5, where occlusions stand out as "anomalous" arcs of one of the intrinsic curves. The detection of occlusions seems easier in this setting, rather than in the "natural" spatial domain of the signals.

### 2.3.2 Incompatible Mapping

The mapping  $\alpha(x)$  may not be compatible with the operators  $P_1, \dots, P_N$ . For instance, affine transformations  $x \rightarrow ax + d$  are not compatible with the operator  $P[l](x) = (l(x), l'(x))$ . In fact, if  $C_l$  is the intrinsic curve generated by  $l(x)$ , the intrinsic curve  $C_r$  generated by  $r(x) = l(ax + d)$  is

$$C_r = \{(p_1, ap_2) : (p_1, p_2) \in C_l\}, \quad (6)$$

a vertically scaled version of  $C_l$ . Affine transformations cause problems also for correlation-based stereo systems [23].

### 2.3.3 Photometric Distortion and Noise

Brightness may vary between images. A simple model that accounts for both geometric and photometric distortion is the following:

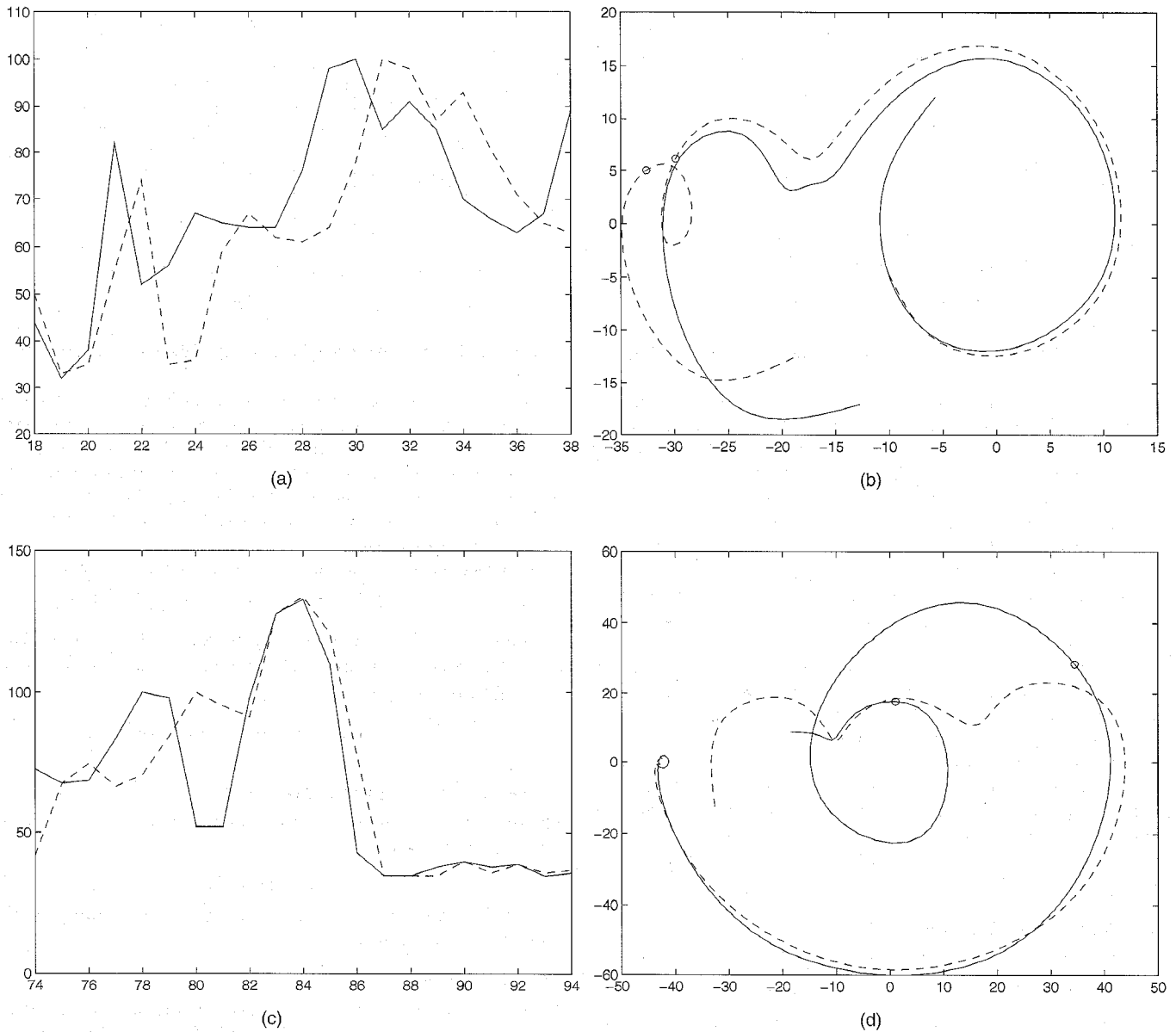


Fig. 5. Details of intensities and intrinsic curves from the image of Fig. 1. (a) Intensity on scanline 92, pixels 18–38. The part of  $r(x)$  (dashed) from pixel 23 to pixel 25 is not matched by  $l(x)$  (solid). (b) Intrinsic curves for (a). The arc of  $C_r$  (dashed) between the two circled points is not matched in  $C_l$  (solid). (c) Intensity on scanline 95, pixels 74–94. The part of  $l(x)$  (solid) from pixel 79 to pixel 81 is not matched by  $r(x)$  (dashed). (d) Intrinsic curves for (c). The arc of  $C_l$  (solid) between the two circled points is not matched in  $C_r$  (dashed).

$$r(x) = A l(ax + d) + B + n(x), \quad (7)$$

where  $A$  and  $B$  represent the difference in contrast and brightness between the two images, and are either constant or varying slowly with respect to  $l(x)$ . The term  $n(x)$  represents “noise,” that is, any discrepancy independent of the signals. The terms  $a$  and  $d$  represent geometric distortion, and, in particular,  $d$  is the disparity we are after.

We have observed empirically that large geometric distortions are less frequent than photometric distortions, in accordance with the results in [1]. Consequently, the terms  $A$  and  $B$  in our model dominate, and we can, therefore, consider their effects alone for the operator  $P[l](x) = (l(x), l'(x))$ . The intrinsic curve  $C_r$  generated by  $r(x) = A l(\alpha(x)) + B$  with compatible  $\alpha(x)$  is

$$C_r = \{(A p_1 + B, A p_2) : (p_1, p_2) \in C_l\}. \quad (8)$$

Hence, transformation (7) induces an isotropic expansion of the curve by a factor  $A$  and a horizontal shift by  $B$ .

### 3 STEREO MATCHING USING INTRINSIC CURVES

In our implementation, we chose the 4D operator  $P(x) = \{l^{(-1)}(x), l(x), l'(x), l''(x)\}$ . The second step is defining a distance between descriptors. The *radial metric* defined in Section 3.1 is a distance between intrinsic-curve points to be used to generate candidate matches. Ambiguities are resolved by maximizing a global quality function by dynamic programming. For efficiency, we first transform the set of candidate matching points into a set of short candidate matching segments via a local aggregation algorithm. Finally, experiments on real stereo pairs prove the effectiveness of our approach.

#### 3.1 Radial Metric

In Section 2.3, we have argued that contrast and brightness bias dominate over geometric distortion and noise. The effects of both brightness bias and noise can be minimized by preprocessing both signals with a zero-mean, low-pass filter. In addition, the zero-

mean filter causes intrinsic curves to loop around the origin.

Contrast differences cause intrinsic curves to expand or contract isotropically. This observation suggests the *radial metric* for intrinsic curves on the plane: Two points have finite distance, equal to their Euclidean distance, when they are collinear with the origin; otherwise, their distance is set to infinity.

For curves in more than two dimensions, we first consider the projection of the curves onto the  $\{l, l'\}$  plane and detect, for each point  $\mathbf{p}$  of the projection of the left curve, the points on the projection of the right curve that are collinear with  $\mathbf{p}$  and with the origin. The distance between such points and  $\mathbf{p}$  is defined as the Euclidean distance of the corresponding points in the 4D space. This simple distance definition allows us to overcome most of the complexities associated with nearest-neighbor computation in higher dimensions [25], [26].

For curves on the plane, we can efficiently determine the candidates  $\{\mathbf{p}_i\}$  by indexing the curve in terms of the phase angle of its points, so that points on the same radial line are stored next to each other. For higher dimensional curves, we compute this same representation for their first two components. With this representation, the complexity of matching is  $O(nk)$ , where  $n$  is the number of samples on  $C_l$ , and  $k$  is the number of loops of  $C_r$  around the origin. For each point in  $C_l$ , we retain the (at most) two closest matching points on  $C_r$ .

With our implementation, we do not need to constrain the maximum disparity for computational efficiency, as most stereo algorithms do, nor do we need to use multirate structures to cope with large disparities. In fact, we work only on intrinsic curves, which have lost track of space: The inverse mapping  $\mathbf{p} \rightarrow x$  is determined only after the matches have been assigned. However, to reduce ambiguity, we divide scanlines into overlapping segments of 120 pixels each in our experiments.

### 3.2 Local Aggregation

For efficiency in the global optimization phase, candidate matching points are first aggregated into candidate matching *segments*. In fact, if two points match, and neighboring points in the two intrinsic curves match with equal or better quality, these two matches should be tied together: In all likelihood, they are either both right or both wrong. This is the intrinsic curve equivalent to the smoothness principle [31], [28].

We, thus, identify a number of candidate match segments as follows. Given two couples of candidate matches  $(\mathbf{p}_{l1}, \mathbf{p}_{r1}), (\mathbf{p}_{l2}, \mathbf{p}_{r2})$ , let  $d_l$  and  $d_r$  be the distances (measured as the arc-length on the intrinsic curve) between  $\mathbf{p}_{l1}$  and  $\mathbf{p}_{l2}$  and between  $\mathbf{p}_{r1}$  and  $\mathbf{p}_{r2}$ , respectively. If  $(d_l - d_r)/d_l$  is within  $[-\delta, \delta]$  for some small  $\delta$ , then we assign  $(\mathbf{p}_{l1}, \mathbf{p}_{r2})$  and  $(\mathbf{p}_{r1}, \mathbf{p}_{r2})$  to the same couple of candidate matching segments. In order to limit the propagation of mismatches, however, the length of the segments (measured as the arc-length of the corresponding segments of intrinsic curves) is kept constant on the left curve  $C_l$ .

A greedy algorithm of complexity  $O(n)$  efficiently constructs a feasible set of candidate matching segments by examining each point of  $C_l$  in scanning order. There are fewer candidate matching segments than candidate matching points, therefore the search space for the optimization is reduced.

### 3.3 Global Optimization

The procedure described in the previous section provides a set of candidate matching segments. We now need a global procedure to pick the correct chain of matches from the pool of candidates. As customary [20], [27], we define an overall quality function to be maximized, together with some constraints to be satisfied. The quality function is

$$Q = \sum_i \exp(-d(\mathbf{p}_{li}, \mathbf{p}_{ri})) \quad (9)$$

where  $d(\mathbf{p}_{li}, \mathbf{p}_{ri})$  is the radial distance between the matching points  $\mathbf{p}_{li}$  and  $\mathbf{p}_{ri}$  in the left and right curves, respectively. The choice of the exponential function is not essential—any other well-behaved function decreasing between 1 and 0 will do (see Fig. 6).

The constraints to be satisfied are [28] *uniqueness* (a point may match at most one point in the other image) and *ordering* (the order of two points in an epipolar line must be the same as the order of their matches in the other image). The latter is violated in rare cases, as in the “double nail illusion.” Two couples of matching segments  $\{c_i = [(l_{i1}, l_{i2}), (r_{i1}, r_{i2})]\}$  and  $\{c_j = [(l_{j1}, l_{j2}), (r_{j1}, r_{j2})]\}$  are *compatible* and *ordered* (that is, they satisfy uniqueness and ordering) if

- 1)  $(l_{i1}, l_{i2})$  and  $(l_{j1}, l_{j2})$  do not overlap, nor do  $(r_{i1}, r_{i2})$  and  $(r_{j1}, r_{j2})$ ;
- 2) the ordering of  $l_{i1}, l_{i2}, l_{j1}, l_{j2}$  is the same as that of  $r_{i1}, r_{i2}, r_{j1}, r_{j2}$ .

Each candidate matching segment is characterized by a quality value  $q$ , which is the sum of the qualities of the matching points that form it.

### 3.4 The Matching Algorithm

Given  $m$  candidate couples of segments  $\{c_i = [(l_{i1}, l_{i2}), (r_{i1}, r_{i2})]\}$ , where  $(l_{i1}, l_{i2})$  and  $(r_{i1}, r_{i2})$  are the endpoints of the  $i$ th segment on the left and right curves, respectively, we now show an algorithm that finds a chain of compatible segments of maximum quality. The quality of a chain is the sum of the qualities  $q_i$  of its matches.

We say that a couple  $c_i$  is a *predecessor* of a couple  $c_j$  (or  $c_j$  is a *successor* to  $c_i$ ) if  $c_i$  and  $c_j$  are compatible and the segments in  $c_i$  are to the left of the segments in  $c_j$ . We denote this relation by  $c_i < c_j$ .

The relation  $c_i < c_j$  is a strict<sup>1</sup> partial ordering, since it is transitive ( $c_i < c_j < c_k \Rightarrow c_i < c_k$ ), asymmetric ( $c_i < c_j \Rightarrow c_j \not< c_i$ ), and irreflexive ( $c_i \not< c_i$ ).

This partial ordering generates a lattice, defined as the directed, acyclic graph in which nodes represent couples, and an edge from  $c_i$  to  $c_j$  means that  $c_i$  is an immediate predecessor of  $c_j$  (that is,  $c_i < c_j$ , and there is no  $k$  such that  $c_i < c_k < c_j$ ). The lattice can be constructed in  $O(m^2)$  time by the algorithm below.

Add dummy couples  $c_0$  and  $c_{m+1}$  that are predecessor and successor to every other couple, respectively. Form the complete partial order graph  $C$ . The nodes of  $C$  are all the couples, and there is an edge between  $c_i$  and  $c_j$  iff  $c_i < c_j$ . This can be done in  $O(m^2)$  by looping over all pairs of couples.

Using Dijkstra's algorithm [29, vol. 1], label each node  $c$  in  $C$  with the length  $L(c)$  of the longest path from  $c_0$  to  $c$ . Although Dijkstra's algorithm is originally a *shortest* path algorithm, the modification necessary to obtain longest paths is trivial. Dijkstra's algorithm takes time  $O(m^2)$ .

Now form the desired lattice by the following procedure.  $\mathcal{L}$  will contain the lattice at the end, and  $\mathcal{F}$  is the frontier set in a breadth-first visit of the graph  $C$ .

- 1:  $\mathcal{L} = \emptyset$
- 2:  $\mathcal{F} = \{c_0\}$
- 3: for  $c \in \mathcal{F}$
- 4:     remove  $c$  from  $\mathcal{F}$
- 5:     create a copy  $C_c$  of the subgraph of  $C$  reachable from  $c$
- 6:     while there are edges  $c \rightarrow c'$  in  $C_c$
- 7:         let  $d = \arg \min_{c'} L(c')$
- 8:         add edge  $c \rightarrow d$  to  $\mathcal{L}$
- 9:         add  $d$  to  $\mathcal{F}$
- 10:     remove from  $C_c$  the subgraph reachable from  $d$
- 11:     end
- 12: end

1. “Strict” here means that a couple cannot be a predecessor to itself.

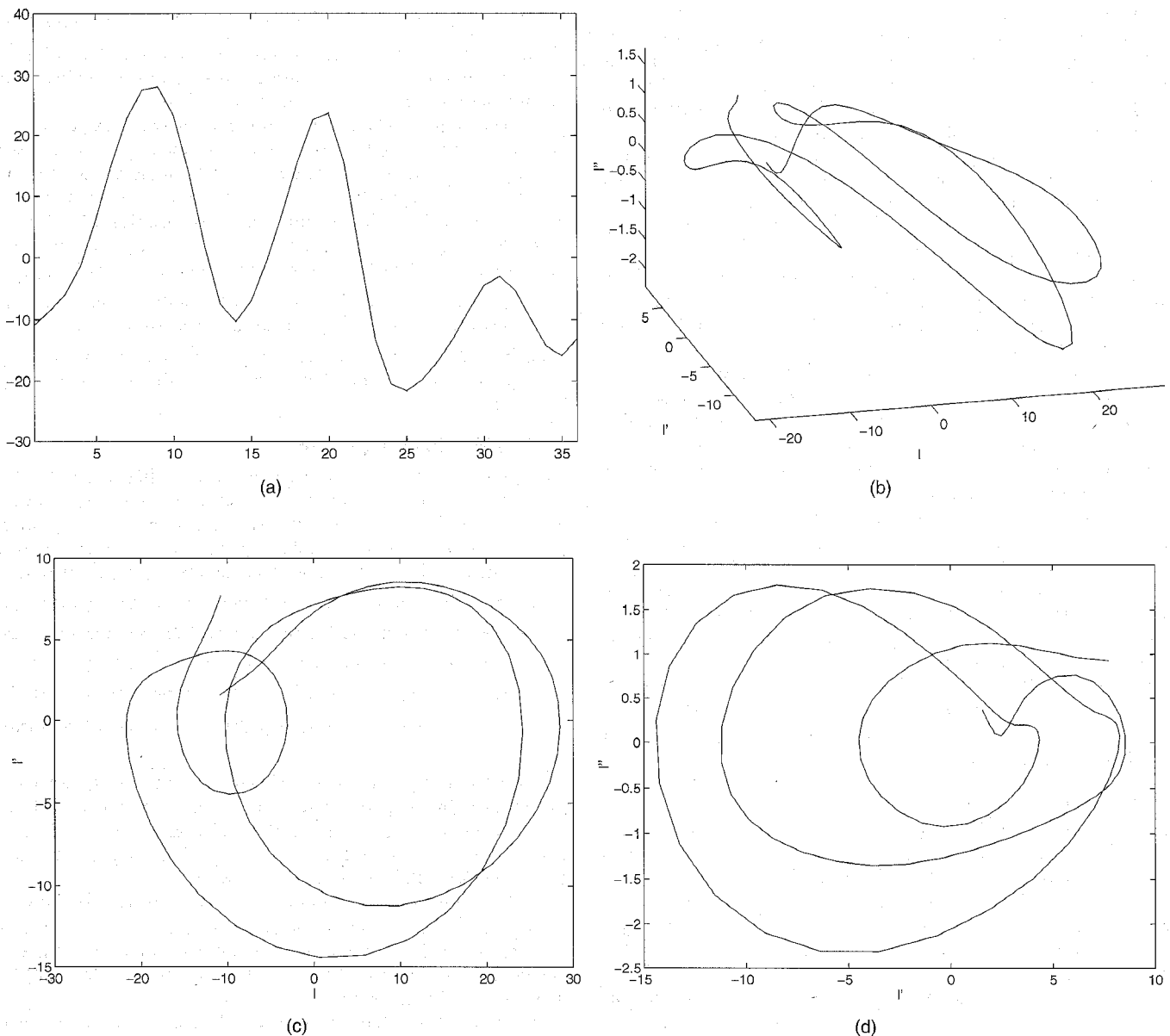


Fig. 6. A signal  $l(x)$  (a), the intrinsic curve generated by  $l(x)$  in the phase space  $(l, l', l'')$  (b), and its projections onto the  $(l, l')$  plane (c) and onto the  $(l', l'')$  plane (d).

In words, the partial order graph  $C$  is traversed breadth-first starting from  $c_0$ , reaching each node by the fastest (line 7) route. The step on line 10 ensures that whenever a node  $e$  is reachable from  $c$  both directly and through  $d$ , the direct connection is not added to the lattice. This connection is redundant, because  $c < e$  can be inferred from  $c < d$  and  $d < e$ .

The work performed by the procedure above for each  $c$  is bounded by the size of the unvisited part of the graph, so the overall complexity is  $O(m^2)$ .

Finally, each node of the lattice can be labeled by its match quality with the convention that  $c_0$  and  $c_{m+1}$  have quality zero. One more application of Dijkstra's algorithm<sup>2</sup> will produce the highest-quality path from  $c_0$  to  $c_{m+1}$ , that is, the globally best match.

### 3.5 Experiments

We have tested our stereo algorithm on the stereo pairs "Clo-

rox" from Stanford University and "Castle" from Carnegie Mellon University (Fig. 7), both subsampled to size  $240 \times 255$  pixels. "Clorox" exhibits a very articulated depth field with many occlusions. "Castle" shows patches with periodically repeated patterns and has high values of disparity (up to about 15 pixels).

The images have been first filtered by a zero-DC notch filter, with an otherwise low-pass frequency response. Each scanline is first divided into overlapping segments 120 pixels long, which reduces the actual disparity search window to approximately  $\pm 50$  pixels, a generous disparity interval in most cases.

The computed disparity maps are represented with pseudocolors. No postprocessing has been performed on the computed disparities. The sparseness of the computed disparity values results from the constant arc-length period sampling procedure (see Fig. 4) and from the fact that candidate segment matches with quality below a threshold have been rejected.

2. Labeling the nodes is equivalent to labeling each edge with the quality of its source node.

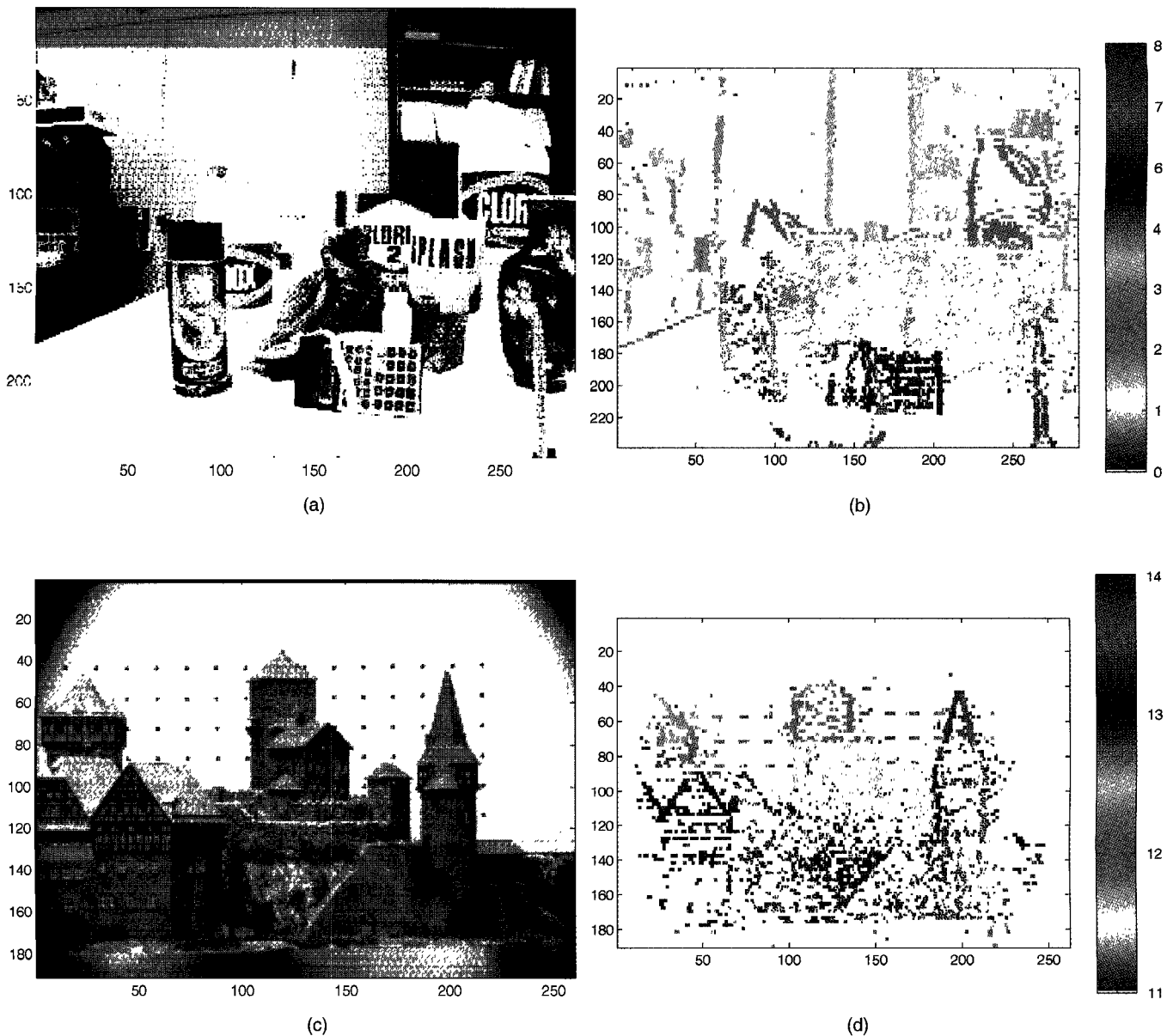


Fig. 7. Experiments with the stereo pairs "Clorox" (a), (b) and "Castle" (c), (d). Only one image is shown for each pair. The computed disparity field is represented with pseudocolors; the correspondence between colors and disparities is given by the color bars on the right.

#### 4 CONCLUSIONS

Intrinsic curves are a new and useful way to think about stereo, and lead to practical matching algorithms. Matching in the space of descriptors makes the amount of disparity irrelevant, and no multiresolution technique is needed even for large displacements. Instead of searching for matches over disparities, one can use efficient algorithms for nearest-neighbor lookup. The concept of compatible mappings elucidates the basic issues in the design of local image descriptors. Extensions to full images are conceptually straightforward but technically challenging. The curves become surfaces or 2D manifolds in more dimensions, but the basis for matching them remains the same.

#### ACKNOWLEDGMENTS

This work was supported by the U.S. National Science Foundation contracts IRI-9509149 and IRI-9506064, Department of Defense MURI contract DAAH04-96-1-0007, and by a grant of the Charles Lee Powell Foundation.

#### REFERENCES

- [1] R.D. Arnold and T.O. Binford, "Geometric Constraints in Stereo Vision," *Proc. SPIE*, vol. 238, pp. 281-292, San Diego, Calif., 1978.
- [2] Y. Ohta and T. Kanade, "Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 139-154, Mar. 1985.
- [3] M.H. Kass, "Computing Stereo Correspondence," master's thesis, Massachusetts Institute of Technology, 1984.
- [4] D.G. Jones and J. Malik, "A Computational Framework for Determining Stereo Correspondence From a Set of Linear Spatial Filters," *Proc. EECV '92*, pp. 395-410, Santa Margherita Ligure, Italy, 1992.
- [5] J. Weng, N. Ahuja, and T.S. Huang, "Matching Two Perspective Views," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 8, pp. 806-825, Aug. 1992.
- [6] A. Blake and C. Marinos, "Shape From Texture: Estimation, Isotropy and Moments," *Artificial Intelligence*, vol. 45, pp. 323-380, 1990.
- [7] B.J. Super and A.C. Bovik, "Shape-From-Texture by Wavelet-Based Measurement of Local Spectral Moments," *Proc. CVPR '92*, pp. 296-301, Champaign, Ill., 1992.

- [8] R. Manmatha, "A Framework for Recovering Affine Transforms Using Points, Lines or Image Brightness," *Proc. CVPR '94*, pp. 141–146, Seattle, Wash., 1994.
- [9] J. Sato and R. Cipolla, "Extracting the Affine Transformation From Texture Moments," *Proc. ECCV '94*, pp. 165–172, Stockholm, 1994.
- [10] J.J. Koenderink and A.J. Van Doorn, "Geometry of Binocular Vision and a Model for Stereopsis," *Biological Cybernetics*, vol. 21, pp. 29–35, 1976.
- [11] K. Kanatani, "Detection of Surface Orientation and Motion From Texture by a Stereological Technique," *Artificial Intelligence*, vol. 23, pp. 213–237, 1984.
- [12] M. Campani and A. Verri, "Motion Analysis From First-Order Properties of Optical Flow," *CVGIP: Image Understanding*, vol. 56, no. 1, pp. 90–107, 1992.
- [13] D.G. Jones and J. Malik, "Determining Three-Dimensional Shape From Orientation and Spatial Frequency Disparities," *Proc. EECV '92*, pp. 661–669, Santa Margherita Ligure, Italy, 1992.
- [14] S.G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.
- [15] J. Malik and P. Perona, "Preattentive Texture Discrimination With Early Vision Mechanisms," *J. Optical Soc. of Am.-A*, vol. 7, no. 5, pp. 923–932, 1990.
- [16] V.I. Arnold, *Ordinary Differential Equations*. Cambridge, Mass.: MIT Press, 1990.
- [17] D.J. Struik, *Lectures on Classical Differential Geometry*. New York: Dover, 1988.
- [18] P.S. Toh and A.K. Forrest, "Occlusion Detection in Early Vision," *Proc. Third Int'l Conf. Computer Vision*, pp. 126–132, Osaka, Japan, 1990.
- [19] J.J. Little and W.E. Gillet, "Direct Evidence for Occlusion in Stereo and Motion," *Proc. ECCV '90*, pp. 336–340, Antibes, France, 1990.
- [20] P.N. Belhumeur and D. Mumford, "A Bayesian Treatment of the Stereo Correspondence Problem Using Half-Occluded Regions," *Proc. CVPR '92*, pp. 506–512, Champaign, Ill., 1992.
- [21] D. Geiger, B. Ladendorf, and A. Yuille, "Occlusions and Binocular Stereo," *Proc. EECV '92*, pp. 425–433, Santa Margherita Ligure, Italy, 1992.
- [22] S.S. Intille and A.F. Bobick, "Disparity-Space Images and Large Occlusion Stereo," *Proc. ECCV '94*, pp. 179–186, Stockholm, 1994.
- [23] Y. Xion and L. Matthies, "Error Analysis of a Real-Time Stereo System," *Proc. CVPR '97*, pp. 1,087–1,093, Puerto Rico, 1997.
- [24] C. Tomasi and R. Manduchi, "Stereo Without Search," Technical Report STAN-CS-TR-95-1543, Stanford Univ., 1995.
- [25] J.K. Uhlmann, "Satisfying General Proximity/Similarity Queries With Metric Trees," *Information Processing Letters*, vol. 40, pp. 175–179, 1991.
- [26] P.N. Yianilos, "Data Structures and Algorithms for Nearest Neighbor Search in General Metric Spaces," *Proc. Fourth ACM-SLAM Symp. Discrete Algorithms*, 1993.
- [27] I.J. Cox, S. Hingorani, B.M. Maggs, and S.B. Rao, "A Maximum Likelihood Stereo Algorithm," *Computer Vision and Image Understanding*, vol. 63, no. 3, pp. 542–567, 1996.
- [28] S.B. Pollard, J.E. Mayhew, and G.P. Frisby, "PMF: A Stereo Correspondence Algorithm Using a Disparity Gradient Limit," *Perception*, vol. 14, pp. 449–470, 1985.
- [29] D.E. Knuth, *The Art of Computer Programming*, 2nd ed. Reading, Mass.: Addison-Wesley, 1973.
- [30] H.H. Baker and T.O. Binford, "Depth From Edge and Intensity Based Stereo," *Proc. Seventh Int'l Joint Conf. Artificial Intelligence*, pp. 631–636, 1981.
- [31] W.E.L. Grimson, "Computational Experiments With a Feature Based Stereo Algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 7, no. 1, pp. 17–34, 1985.

## Computing Joint Distributions of 2D Moving Median Filters With Applications to Detection of Edges

Dawei Huang, *Member, IEEE*, and William T.M. Dunsmuir

**Abstract**—This paper derives the joint distribution of medians over moving windows in a two dimensional noisy image. The general formulation presented allows derivation of the probability distribution, needed to evaluate the probability of failing to detect an edge when present ("edge miss probability") and the probability of falsely detecting a nonexistent edge.

**Index Terms**—Image analysis, moving medians, joint distributions, edge detection, edge miss probability.

### 1 INTRODUCTION

MEDIAN filters are now widely used in image processing because they are effective at smoothing out noise while preserving various important root signals which frequently occur in typical two dimensional images. There are many types of median filters which can be considered. In particular, the choice of pixel window over which the filter is defined provides a variety of filters for different applications. In this paper, we will consider pixel windows defined on the square, containing an odd number of pixels, on the X-shaped window, defined only on the diagonal elements of the square, and, on the cross-shaped window, which uses pixels aligned vertically and horizontally. In [1], these are defined as SQUARE( $2L + 1$ ), CROSS( $2L + 1$ ), and XSHAPE( $2L + 1$ ). These shapes provide differing abilities for preserving vertically, horizontally, or diagonally aligned edges in an image. Additional flexibility can be introduced to median filters through use of nonuniform weights at different pixels in the window. For example, exponentially decaying weights can be applied as in [3]. In this paper, we will consider exclusively the case of equally weighted medians, in which case the median filter is defined simply as the median of the data observed at the pixels in the window.

When prefiltering a two-dimensional image by moving median filters of particular interest is the probability of failing to detect a real edge in the original, noise-free image ("edge miss probability"). Of equal interest is the probability of falsely concluding that an edge exists (detecting a "pseudoedge"). These probabilities can be obtained from the joint distribution of medians calculated from data in two overlapping windows centered one pixel apart. Derivation of this distribution is complicated by the fact that the data in each of the two windows is not sampled from the same distribution. The derivation of this joint distribution was considered to be "an extremely complex problem" (see [1]). This paper solves this problem by providing a derivation of the required distributions in a series of theorems—see Section 2.

Let  $(i, j)$  be the coordinates of a pixel, where  $i$  refers to the rows (vertical direction) and  $j$  references to the columns (hori-

- D. Huang is with the School of Mathematics, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia.  
E-mail: huang@fsc.qut.edu.au.
- W.T.M. Dunsmuir is with the School of Mathematics, University of New South Wales, Sydney, NSW 2052, Australia.  
E-mail: W.Dunsmuir@unsw.edu.au.

Manuscript received 19 Dec. 1996. Recommended for acceptance by J. Daugman. For information on obtaining reprints of this article, please send e-mail to: [tpami@computer.org](mailto:tpami@computer.org), and reference IEEECS Log Number 106154.