Stereo Without Search

Carlo Tomasi
Computer Science Department
Stanford University
Stanford, CA 94305
tomasi@cs.stanford.edu

Roberto Manduchi
Computer Science Division
University of California at Berkeley
Berkeley, CA 94720
manduchi@nefeli.eecs.berkeley.edu

Abstract

Search is not inherent in the correspondence problem. We propose a representation of images, called *intrinsic curves*, that combines the ideas of associative storage of images with connectedness of the representation: intrinsic curves are the paths that a set of local image descriptors trace as an image scanline is traversed from left to right. Curves become surfaces when full images are considered instead of scanlines. Because only the path in the space of descriptors is used for matching, intrinsic curves loose track of space, and are invariant with respect to disparity under ideal circumstances. Establishing stereo correspondences then becomes a trivial lookup problem. We also show how to use intrinsic curves to match real images in the presence of noise, brightness bias, contrast fluctuations, and moderate geometric distortion, and we show how intrinsic curves can be used to deal with image ambiguity and occlusions. We carry out experiments on single-scanline matching to prove the feasibility of the approach and illustrate its main features.

This research was supported by the National Science Foundation under contract IRI-9496205.

Contents

1 Introduction

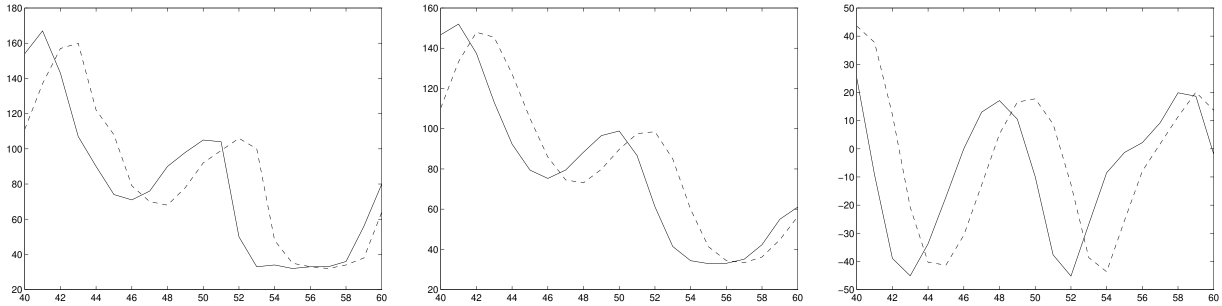
The computation of stereo correspondences has traditionally been associated with search: for every point in the left image the right image is searched for a similar point. In this paper we show that search is not inherent in the correspondence problem. The way out of search is associative memory, and essentially inverts the way images are represented. Rather than storing image intensities by their position in the image, the usual array $I(\mathbf{x})$, we can store image positions by their appearance: in a sense, $\mathbf{x}(I)$. 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: ambiguity and disguise. Ambiguity means that different image points can look the same, so memory locations can be crowded. Disguise occurs when corresponding points in the two images look different because of the viewpoint change or of image noise. In this case, points that should go in the same memory location do not. We address ambiguity by using a richer description of point $\mathbf{x}$ than just its intensity I. We deal with disguise by analyzing possible changes between images. This analysis lets us identify occlusions, and tells us where to look next if a memory location is missing a point. In addition, our description of image appearance varies continuously with $\mathbf{x}$, so points that are close in the images are also close in memory. In other words, while the memory addresses are not computed from $\mathbf{x}$, they vary continuously with $\mathbf{x}$, so they preserve contiguity in the image.

In this paper we illustrate this approach by matching corresponding scanlines. Matching entire images is conceptually simpler, because it is easier to find rich image descriptions, but it is technically more complex. It also makes it harder to draw plots, so the explanation would be obscured. We discuss the necessary modifications for entire images in section 5. Furthermore, we restrict our discussion to simple image descriptors: two numbers per image point. Again, richer descriptors simplify the problem because they reduce ambiguity, but plots become more difficult to draw.

Here is our representation of, say, the left scanline (solid lines in figure 2(a)). A lowpass filtered version of the image intensity $l(x)$ and its derivative $l'(x)$ are computed everywhere (solid lines in figure 2 (b) and (c)) and are plotted against each other (solid lines in figures 3 (a) and (b)). When plotting l' versus l we lose track of space, that is, of the coordinate x which merely parameterizes the curve $l'(l)$. Of course, this parameter is stored, but it plays no role in the shape of the curve. If $l(x)$ is replaced by a shifted replica $r(x) = l(x + d)$, the curve of figure 3 (b) remains the same. Because of this invariance to displacements, we call the curve of figure 3 (b) an *intrinsic curve*. More general geometric transformations $r(x) = l(\alpha(x))$ between l and r can deform an intrinsic curve, but the deformations can be predicted as discussed in section 3. The dashed curves in figures 2 and 3 show the construction of the intrinsic curve for the scanline $r(x)$ taken from a different viewing position. Matching the two intrinsic curves in figure 3 is a nearest-neighbor problem, as shown in section 3.

Ambiguities cause intrinsic curves to self-intersect. Clearly, the richer the description is, that is, the higher the dimensionality in which an intrinsic curve lives, the less likely self-intersections are. For instance, if we use three descriptors per point, rather than two, the curve lives in a three-dimensional space. A curve in space is much less likely to intersect itself than a curve on the plane.

Figure 1: Test image "Trees" from SRI - frame 1.

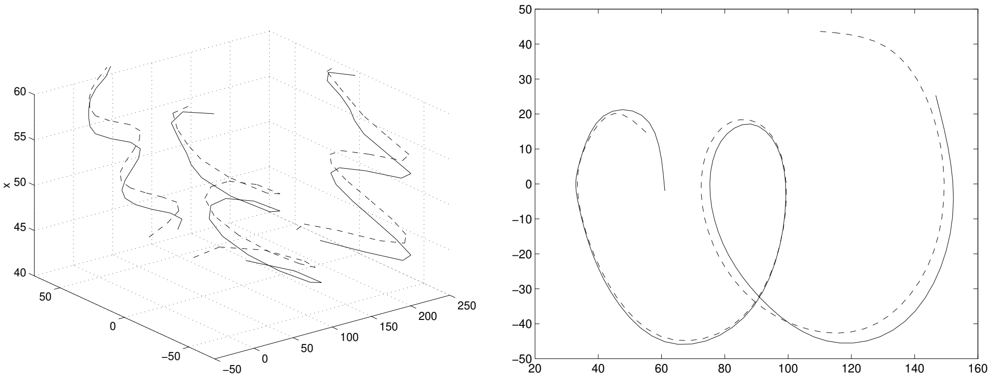


(a) (b) (c)

Figure 2: Scanline 68 from the image of figure 1, pixels 40–60 (solid line: frame 1, dashed line: frame 2). (a) Intensity, (b) its lowpass filtered version, and (c) its derivative.

Occlusions are pieces of one curve that have no corresponding piece in the other (see figures 13 and 14). We show in section 4 that occlusions produce new loops in intrinsic curves and consequently stand out very clearly.

Thus, stereo matching can proceed by "stapling together" intrinsic curves from the two images. Corresponding points are nearest neighbors, and occlusions are left over as unmatched loops. Having to "look for" the corresponding point on the other curve during stapling seems to imply that search has crept back into our solution to stereo. In section 3, however, we show that this is not so. In fact, a nearest neighbor lookup is a much easier problem than search. In addition, we know what direction to look in. Finally, the nearest neighbor problem arises for reasons that are entirely different from those that caused search in the traditional approaches to stereo. Search in these approaches was necessary because of image displacements, that is, because of variations in the *domain* of the function $l(x)$. Nearest neighbor matches are now necessary because of disguise, that is, changes of appearance. In our approach these have to do mostly with the *range* of $l(x)$. The relation of our approach with matching techniques based on Sum of Squared Differences (SSD) correlation is discussed in appendix A.



(a) (b)

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

2 Intrinsic Curves

An efficient procedure for matching two signals is to consider a vectorial description of the local intensity variation at every point. Then two points from the two images are match candidates if the local descriptions are "close" to each other. A similar idea is at the basis of the stereo algorithms of Kass [Kas84] and of Jones and Malik [JM92a]. The local description of a signal at a given point is the vector composed by the outputs of a bank of operators (in the case of [Kas84] and [JM92a], linear filters) at the same point. Multiple attributes (intensity, edgeness and cornerness) are also considered in the approach of Weng, Ahuja and Huang [WAH92]. In this section we define intrinsic curves more generally. We also identify the geometric mappings

$$r(x) = l(\alpha(x)) \tag{1}$$

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

Definition of an *intrinsic curve*. Suppose that the N operators $P_1, \ldots, P_N$ are applied to the intensity signal $l(x)$ to produce the new signals

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

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

$$\mathbf{p}(x) = (p_1(x), \ldots, p_N(x)) \tag{2}$$

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

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

$\mathcal{C}$ is called the *intrinsic curve* generated by $l(x)$ through the operators $P_1, \ldots, P_N$.

An example of intrinsic curve construction is shown in figures 2, 3.

2.1 Compatible Mappings

An intrinsic curve $\mathcal{C}$ does not characterize its generator signal $l(x)$. In fact, to reconstruct $l(x)$ from $\mathcal{C}$ we need (i) to know the parametrization (3) and (ii) to invert the operators P_n. The transformation from $l(x)$ to $\mathcal{C}$ has, in a sense, lost track of space, and the space coordinate x is now but one of infinitely many parametrizations of $\mathcal{C}$. Furthermore, the operators P_n may not be invertible.

While any reparametrization of $\mathcal{C}$ leaves $\mathcal{C}$ unchanged, reparametrizing $l(x)$ to $l(\alpha(x))$ can in general modify $\mathcal{C}$. For instance, if where the prime denotes differentiation, the new components of $\mathcal{C}$ 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 $\tilde{\mathbf{p}}(x)$ traces a new curve that is modulated by $\alpha'(x)$ in its second component.

Definition of a *compatible mapping.* A mapping $x \rightarrow \alpha(x)$ is said to be *compatible* with the operators $P_1, \ldots, P_N$ if for any 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 $P_1, \ldots, P_N$.

2.1.1 Examples

Constant Displacement. Let the operators P_n in (2) be shift-invariant:

$$l(x) \rightarrow l(x+d) \Rightarrow p_n(x) \rightarrow p_n(x+d) \ .$$

The constant displacements

$$\alpha(x) = x + d$$

are compatible with shift-invariant operators.

Affine Mapping. The affine mappings of the form

$$\alpha(x) = ax + d$$

are compatible with the operators

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

defined wherever $\frac{\mathrm{d}^{n+1}}{\mathrm{d}x^{n+1}} l(x) \neq 0$. This is proved immediately by noting that

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n}[l(ax+d)] = a^n [\frac{\mathrm{d}^n l}{\mathrm{d}x^n}] l(ax+d) \ .$$

Semi-Commutative Mapping. If the mapping $x \to \alpha(x)$ is regarded as an operator A applied to $l(x)$,

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

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

$$P_n A = D P_n$$

where D is a diffeomorphism independent of n. In fact, in this case, the mapping $x \to \alpha(x)$ maps every $p_n(x)$ to $[DP_n](x)$, so the intrinsic curve is simply reparameterized. Both previous examples are special cases of semi-commutative mappings.

Thus, intrinsic curves can be regarded as invariants with respect to the set of compatible mappings, and provide a more general description than "classical" invariants such as function moments [BM90],[SB92],[Man94],[SC94]. Affine mappings are a popular model for the transformation between the two images of a stereo pair [KVD76],[Kan84],[CV92], and shift-invariant filters are often used for image descriptors [Kas84],[JM92a],[JM92b],[Mal89],[MP90]. The fact that in general affine mappings are not compatible with shift-invariant operators is therefore important. This was pointed out in [Kas84], where a clever analysis of the effect of filtering a signal undergoing affine geometrical distorsion is carried out. From the results of [Kas84], we can assume that the intrinsic curves are approximately invariant with diffeomorphisms $x \to \alpha(x)$, so long as the supports of the filters' kernels are narrow and $\alpha(x)$ is close to the identity function. In the remainder of this section we assume that the mapping $\alpha(x)$ is a diffeomorphism (which, in particular, implies that it is monotone and continuous). The case of discontinuous or not one-to-one $\alpha(x)$ will be considered in section 4, where we treat occlusions. In addition, we assume throughout this paper that both the input signals $l(x)$, $r(x)$ and the operators P_n are continuous, so that the intrinsic curves are connected.

If the transformation between left and right image were just a mapping $\alpha(x)$ compatible with the operators $P_1, \ldots, P_N$, stereo matching would be nearly trivial. In fact, to determine $\alpha(x)$ from the observation of $l(x)$ and of $r(x) = l(\alpha(x))$, intrinsic curve are first computed from the two signals. For each signal, the parametrization (3) is stored, so that every point on either curve can be traced back to its image coordinate x via table lookup. Because of compatibility, the two intrinsic curves coincide. For every point $\mathbf{p}$ that belongs to both of them, the corresponding image coordinates are a match, with the sole exception of points where the intrinsic curves self-intersect. Self-intersections are treated in chapter 3 below.

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* [Kas84] $\phi(l(x))$ combine with *geometric distortions* [Kas84] $l(\alpha(x))$, so that the range of l is affected in addition to its domain. Third, noise corrupts both $l(x)$ and $r(x)$. Fourth, macroscopic phenomena such as occlusions and specularities must be dealt with in stereo matching. Photometric distortions and noise are discussed in section 3, while section 4 shows how occlusions can be effectively detected by using intrinsic curves.

In the remainder of this section, however, we confine ourselves to single intrinsic curves, and we show that not every curve in R^N can be an intrinsic curve. In particular, for $N = 2$ and with

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

intrinsic curves must satisfy quite a few geometrical an topological properties, as shown in the next subsection.

2.2 Geometrical and Topological Properties of Intrinsic Curves

Our definition of intrinsic curves is quite general. Their properties depend on the characteristics of the operators P_n in (2). In this section, we concentrate on the case $N = 2$ with the following choice for these operators:

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

where $l'(x)$ is the derivative of l with respect to x. Vector $\mathbf{p}(x)$ is thus composed by the first two terms of the Taylor expansion of $l(x)$ around x, and each point on the intrinsic curve generated by $l(x)$ represents a description of the local behavior of $l(x)$. This choice satisfies the criteria of richness, compactness and sensitivity discussed in [Kas84], but is not stable, as the differentiation is sensitive to noise. To overcome this problem, in the practical implementation of the algorithm we will first low-pass filter the input signal to reduce noise. However, in this section we ignore the presence of this filter and assume that its effect on the signal is small. In other words, we assume that the frequency response of the filter is flat on most of the spectral support of the signal.

With the choice (4) of operators, intrinsic curves are defined on a plane, reminiscent of the *phase space* of systems theory [Arn90]. The abscissa of a point $\mathbf{p}(x)$ corresponds to a value $l(x)$, while the ordinate corresponds to $l'(x)$. Note that the spatial coordinate x is lost when l' is plotted versus l. We assume for now that the first and second derivative of $l(x)$ exist for every x, so that the curves do not have cuspidal points. Let $l(x)$ generate curve $\mathcal{C}$, and assume that a given point $\mathbf{p}(x)$ is traversed just once. We can define an orientation at $\mathbf{p}$ by computing the unit-norm tangent $\mathbf{t}(\mathbf{p})$ of the curve:

$$\mathbf{t}(\mathbf{p}) = \frac{(l'(x), l''(x))}{\sqrt{(l'(x))^2 + (l''(x))^2}} \, . \tag{5}$$

The values of $\mathbf{t}(\mathbf{p})$ depend on the position of $\mathbf{p}$ as follows:

- If $\mathbf{p}$ lies on the upper open half-plane, where $l'(x) > 0$, $\mathbf{t}(\mathbf{p})$ assumes values in the right open half-circle $\{(p_1, p_2); p_1^2 + p_2^2 = 1; p_1 > 0\}$. When $\mathbf{p}$ lies on the lower open half-plane, $\mathbf{t}(\mathbf{p})$ is in the left open half-circle.

- If $\mathbf{p}$ lies on the axis of the abscissas, where $l'(x) = 0$, then $\mathbf{t}(\mathbf{p}) = (0, +-1)$. In other words, when crossing the axis of the abscissas, the curve tangent forms an angle of $\pm\pi/2$ with it.

These rules on intrinsic curves are schematically represented in figure 4. Note that if $l'(x) = l''(x) = 0$ (e.g., in a segment where the signal is constant), $\mathbf{p}$ is singular with respect to x [Str88]. In such a case, the tangent can be defined by continuity. On the other hand, an intrinsic curve can be singular only on the axis of the abscissas.

¿From these rules it follows that *intrinsic curves are naturally oriented clockwise*: they are traversed left-to-right in the upper half-plane and right-to-left in the lower. Furthermore, *any loop must intersect the axis of the abscissas*. We will have more to say about loops in the sections to come.

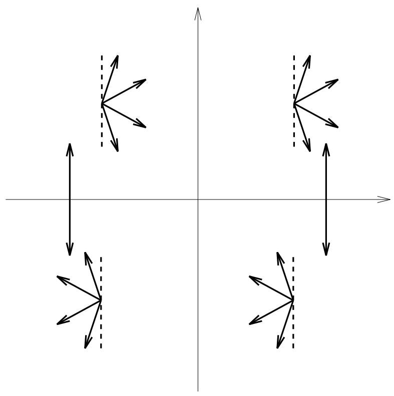


Figure 4: The allowed directions for the tangent of an intrinsic curve $[l, l']$. The dashed segments represent the boundaries of the admissible regions for the tangent.

3 Stapling Intrinsic Curves Together

We have seen in section 2 that two signals $l(x)$ and $r(x)$ related by a compatible mapping generate the same intrinsic curve, and stereo matching becomes a trivial lookup problem. Unfortunately, the hypotheses that lead to this simple situation are seldom satisfied in practice. In this section we first discuss the main deviations from the ideal case and propose a first-order model for the discrepancies. This model accounts for stereo disparity, image shrinkage or dilation, brightness bias, contrast amplification, and noise. We show that both brightness bias and noise can be reduced by filtering, while stereo disparity is exactly accounted for by intrinsic curves. We then argue that contrast amplification dominates the remaining deviations, and we describe an intrinsic curve-matching method tuned to cope with this problem.

Both inaccuracies and gross errors can occur during any matching procedure. For inaccuracies, we first show that intrinsic curves lead naturally to a resampling of the images that is denser where matching is more reliable. Then, we analyze the disparity errors related to image shrinkage or dilation and to a residual brightness bias that may be left even after prefiltering. Gross errors, on the other hand, are caused by ambiguous image regions that look similar to each other. We propose a method for addressing ambiguity which, contrary to most previous approaches, does not rely on disparity values, and can therefore be applied *before* any disparity is computed.

3.1 Deviations from the Ideal Case

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

$$r(x) = l(\alpha(x)) \tag{6}$$

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

No mapping. In certain cases, $\alpha(x)$ may not even exist, such as when regions of $l(x)$ or $r(x)$ are occluded. Occlusions are treated in section 4.

Incompatible mapping. The mapping $x \rightarrow \alpha(x)$ that relates $l(x)$ and $r(x)$ as in (6) is not compatible with respect to the operators $P_1, \ldots, P_N$ that are used to build the intrinsic curves. For instance, affine transformations $x \rightarrow ax + d$ are not compatible with the operators defined in section 2.2, $\mathbf{p}(x) = (l(x), l'(x))$. In fact, if $\mathcal{L}$ is the intrinsic curve generated by $l(x)$, the intrinsic curve $\mathcal{R}$ generated by $r(x) = l(ax + d)$ is

$$\mathcal{R} = \{(p_1, ap_2) : (p_1, p_2) \in \mathcal{L}\} \tag{7}$$

which is a vertically expanded ($a > 1$) or compressed ($a < 1$) version of $\mathcal{L}$.

Photometric distortion and noise. The constant-brightness hypothesis implied by relation (6) is not satisfied. A convenient model that accounts for both geometric and photometric distortion is the following (see [HS92] for a general discussion of related issues):

$$r(x) = B + Cl(ax + d) + n(x) \ . \tag{8}$$

In this model, B and C represent the difference in brightness and contrast between the two images, and are either constant or varying slowly with respect to the dynamics of the signal. 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 inter-frame disparity we are after.

Let us consider the effects of B and C alone (that is, assume $n(x) = 0$ and a compatible $\alpha(x)$). The intrinsic curve $\mathcal{R}$ generated by $r(x) = B + Cl(\alpha(x))$ with compatible $\alpha(x)$ is

$$\mathcal{R} = \{(Cp_1 + B, Cp_2) : (p_1, p_2) \in \mathcal{L}\} \ . \tag{9}$$

Hence, transformation (8) induces an isotropic expansion of the curve by a factor C and a displacement by B along the horizontal direction. Note that *the tangent at corresponding points of the two curves is the same.*

The effects of both brightness bias and noise can be neutralized by preprocessing both signals with a zero-mean filter with an otherwise lowpass frequency response. In order to reduce the effects of the multiplicative term C, Kass [Kas84] suggested to work on the logarithm of the intensity. Then, a second linear filter would be capable of suppressing the difference of contrast. This idea underlies the theory of homomorphic signal processing [OS75]. However, if we have already extracted the mean from the signals in order to reduce the effects of B, then we have to compute the (complex) logarithm of negative values, and address the problem of "phase unwrapping" [OS75].

We choose a different route. In fact, after testing several real-world images, we have observed that the shape of intrinsic curves is altered mostly after photometric distortions, for example as a consequence of the different viewing position of the two cameras, optical attenuation and sensitivity of the image sensors [VP87], [Kas84], [WAH92]. Large geometric distortions that give raise to vertical dilation or shrinking of the intrinsic curve are less likely to happen than photometric distorsions. Such an observation is in accordance with the results of Arnold and Binford [AB78]. Consequently, we believe that the terms B and C in our model are dominant over the geometric distortion related to a.

We therefore design our algorithm to "staple" together the two curves (that is, to determine a pointwise correspondence between $\mathcal{L}$ and $\mathcal{R}$) so that it does not suffer from contrast difference. We

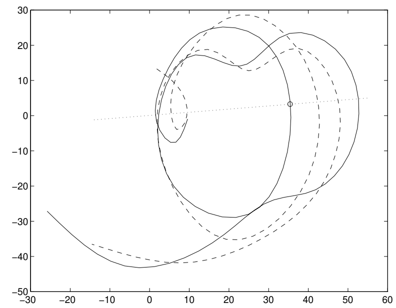


Figure 5: The radial line passing through point $\mathbf{p}$ (circled) in $\mathcal{L}$(solid line). The match candidates are the intersection of the radial line with curve $\mathcal{R}$ (dashed line).

thereby avoid the need for homomorphic preprocessing. This leaves image dilation or shrinking (modeled by the term a in equation (8)) and a possible leftover intensity bias as the only terms of our model (8) that have not been accounted for by our method. We analyze their effects in section 3.3. More macroscopic discrepancies, related to image ambiguity and occlusions, are addressed in sections 3.4 and 4.

3.2 Stapling: Nearest Neighbor Along the Radial Line

In this section we describe the basic idea of our algorithm to "staple together" two curves. In the previous discussion, we concluded that brightness bias and noise can be reduced by prefiltering. Thus, the most important cause of discrepancy, other than occlusions and ambiguity, is the brightness amplification term C in our model (8). The "stapling" procedure described in this section accounts for this amplification exactly.

Assume that the geometric distorsion between signal $l(x)$ and signal $r(x)$ is compatible. A multiplicative photometric constant C induces isotropic expansion of the intrinsic curve. Consequently, a point $\mathbf{p}$ of the intrinsic curve $\mathcal{L}$ generated by $l(x)$ corresponds to a point $\mathbf{p}_{corr}$ of the intrinsic curve $\mathcal{R}$ generated by $r(x)$, which is collinear with $\mathbf{p}$ and the origin. This suggests the following algorithm: choose the candidate matches to $\mathbf{p}$ among the intersections of line $\{s\mathbf{p}, s \in R\}$ with $\mathcal{R}$. We will call such a line the *radial line*. An example of radial line is shown in figure 5.

Other criteria have been employed in [Kas84] and in [JM92a] to determine the "closest" match. The algorithm of Kass chooses the points on $\mathcal{R}$ which are closest in the L_2 norm to $\mathbf{p}$, while Malik and Jones looks for the closest points in the L_1 norm. We can visualize such procedures through our intrinsic curves. In the first case, one would expand a circle around $\mathbf{p}$ until it bumps into $\mathcal{R}$, while in the second case the figure being expanded would be a diamond. Our previous arguments prove that, at least where the difference of contrast is the principal source of distortion, our technique provides the most accurate match.

It is important to realize that "searching" along the radial line is essentially different from "searching" in conventional signal matching algorithm — this is why we use term "stapling" instead of "search" here. In fact, in an ideal situation, no search is required, as the two intrinsic curves are identical. The need for stapling stems from departures from the ideal case due to photometric and affine geometric distorsions. On the other hand, traditional stereo matching algorithms have to perform a search even in the ideal case.

Looking for a match on the radial line is a nearest neighbor problem. In our implementation, intrinsic curves are represented as cubic splines, and nearest neighbors are found by looking them up among the elements of a fine sampling of the splines. A more efficient and precise (but more elaborate) algorithm can be written directly in terms of the polynomial spline segments by using standard techniques of computational geometry. More specifically, the endpoints of the splines representing each curve are stored into a sweep list (with a sweeping half-line being rotated around the origin). Setting up this list requires $O(n \log n)$ time. The set of intersections of a given line through the origin with each curve can then be found by binary search in $O(\log n)$ time.[1] In addition, the practical time complexity of this algorithm can be reduced by using hashing techniques, based on the fact that the two intrinsic curves can be expected to be close to each other, and continuation methods, relying on the continuity of the intrinsic curves.

Regardless of the algorithm used to determine the nearest neighbors, the uniform sampling of intrinsic curves yields a representation of images that emphasizes "busy" regions, where matching can be expected to be more reliable than in "dull" regions. To see this, assume that the trajectory $\mathbf{p}(x)$ goes through the points of the segment $\mathcal{C}(\mathbf{p}_1, \mathbf{p}_2)$ of curve delimited by $\mathbf{p}_1$ and $\mathbf{p}_2$ just once, and let $\mathbf{p}(x_1) = \mathbf{p}_1, \mathbf{p}(x_2) = \mathbf{p}_2$. Then the arc length of the segment is given by

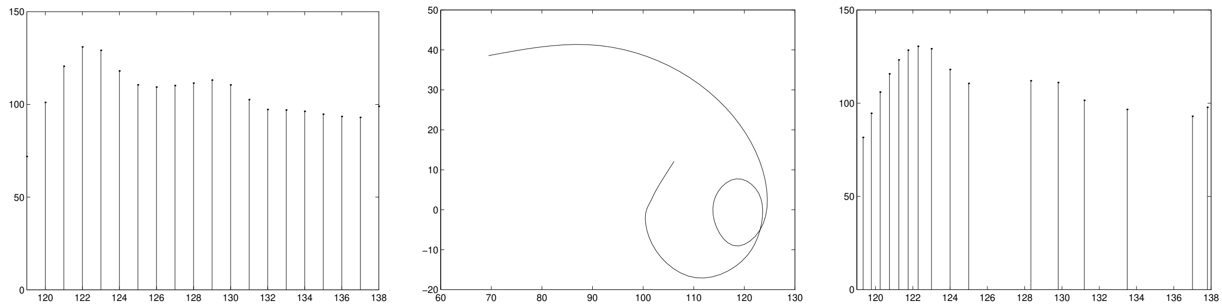
$$\text{arc length } \mathcal{C}(\mathbf{p}_1, \mathbf{p}_2) = \int_{x_1}^{x_2} \sqrt{(l'(x))^2 + (l''(x))^2} \mathrm{d}x \ . \tag{10}$$

Thus for a given distance $|x_2 - x_1|$ in the image we expect longer arc lengths if the curve lies far from the axis of the abscissas (i.e., if $l'(x)$ is large). Conversely, a segment of the curve with a given arc length represents a segment of the signal $l(x)$ which is shorter as the curve lies farther from the axis of the abscissas.

This observation suggests a sort of "adaptive" sampling paradigm for $l(x)$. Assume to sample the curve $\mathcal{C}$ at constant-width intervals, that is, by keeping the arc length of the segments $\mathcal{C}(\mathbf{p}_i, \mathbf{p}_{i+1})$ constant. This procedure corresponds to sampling signal $l(x)$ on a nonuniform grid: the grid will be less dense in areas characterized by small values of $l'(x)$ and $l''(x)$ (where the signal is "flat"), and denser if $l'(x)$ and $l''(x)$ are larger (where the signal "busyness" is higher). This looks like a useful sampling strategy for signal matching. In fact, it is well known (see e.g. [HS92]) that a match is expected to be less robust (with respect, for example, to noise and to quantization errors) in regions where the signal is "flat". The adaptive sampling procedure leads to concentrating estimates in reliable areas. As an example, consider the signal of figure 6 (a), corresponding to the lowpass filtered version of a scanline segment from figure 1. The corresponding intrinsic curve is shown in figure 6(b). If we sample the curve with constant arclength period, we induce a nonuniform sampling of the signal, as shown in figure 6 (c). Samples are denser where the signal busyness is higher, less dense where the signal is flat.

We can extend relation (10) to curve segments containing loops by parametrizing the curve with the arc length s defined by $\mathbf{p}(s) = \mathcal{C}(\mathbf{p}_0, \mathbf{p})$ for some starting point $\mathbf{p}_0$. Note that s is monotone with x: this sense of traversal is the only "memory" left of space.

[1] We thank Leonidas Guibas for suggesting this algorithm.



(a) (b) (c)

Figure 6: A signal sampled on a uniform grid (a) and on an nonuniform grid (c) induced by the uniform arclength sampling of the intrinsic curve (b).

3.3 Matching Inaccuracies

The stapling technique described in the previous section accounts for differences in contrast (image amplification) between the left and the right image. However, if other kinds of photometric distorsion and noise are present, our match will suffer from errors. Match errors typically fall into one of the following categories:

1. *Inaccuracy*: we match a point x_0 of signal $l(x)$ with a point belonging to a "small" neighborhood of the corresponding point of $r(x)$.

2. *Mismatch*: we match a point x_0 of signal $l(x)$ with a point x_1 of $r(x)$ which is entirely incorrect, but such that the "local description" of $r(x)$ around x_1 is similar to the local description of $l(x)$ around x_0. Mismatch is equivalent to *local ambiguity*, which must be resolved at a higher level.

For example, in the case of correlation-based algorithms, inaccuracy corresponds to small deviations of the correlation surface peak, while mismatch corresponds to wrong peak selection [BAR93]. Similar considerations apply to stapling. Noise may change the shape of curve $\mathcal{R}$ in such a way that the stapling is not accurate, or may induce new potentially dangerous match candidates, i.e., increase ambiguity. In certain cases, the radial line may not even cross $\mathcal{R}$. We address the problem of the correct choice among matching candidates in section 3.4. In this section, we analyze the effect of photometric and geometric distorsion on the match accuracy in a simple example.

Let

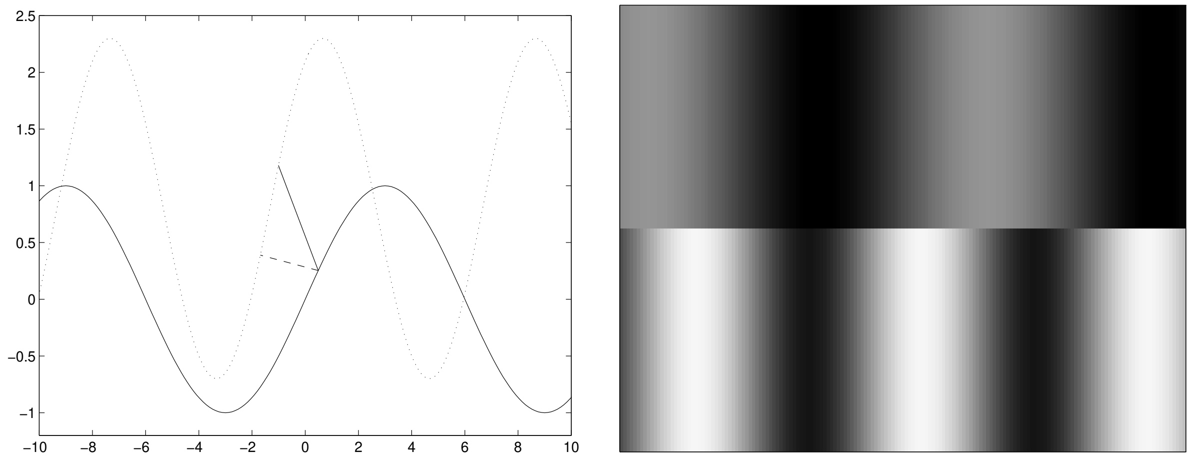
$$l(x) = \sin \omega x \ , \ r(x) = B + A \sin(\omega(ax + d)) \tag{11}$$

with $B \neq 0$ and $A, a \neq 1$ (figure 7). Signals $l(x)$ and $r(x)$ generate the intrinsic curves (ellipses, figure 8)

$$\mathcal{L} = \{(p_1(x), p_2(x)) : p_1(x) = \sin \omega x, p_2(x) = \omega \cos \omega x\} \tag{12}$$

$$\mathcal{R} = \{(p_1(x), p_2(x)) : p_1(x) = B + A \sin(\omega(ax + d)), p_2(x) = Aa\omega \cos(\omega(ax + d))\} \ . \tag{13}$$

Signals $l(x)$ and $r(x)$ are periodic. Therefore, the intrinsic curves they generate are closed, and infinitely many image points x map into the same curve point $\mathbf{p}$. To overcome this ambiguity, we



(a) (b)

Figure 7: (a) Signal $l(x)$ (solid line) and $r(x)$ (dotted line) of our example. The correct match is represented by the solid segment, the estimated match by the dashed segment. (b) Images that generate $l(x)$ (upper half) and $r(x)$ (lower half).

restrict our attention to points belonging to a period of the sinusoids including $x = 0$, so that the correspondence becomes one-to-one. Namely, point $\mathbf{p} = (p_1, p_2)$ of $\mathcal{L}$ corresponds, for $l(x)$, to

$$x = \frac{\arcsin(p_1)}{\omega} \tag{14}$$

while point $\mathbf{p} = (p_1, p_2)$ of $\mathcal{R}$ corresponds, for $r(x)$, to

$$x = \frac{\arcsin\left(\frac{p_1 - B}{A}\right) - \omega d}{\omega a} \tag{15}$$

Relation (11) maps point x relative to $l(x)$ into the "correct" point

$$x_{corr} = (x - d)/a \tag{16}$$

relative to $r(x)$. For a given point $\mathbf{p}$ of $\mathcal{L}$, corresponding to a certain x via (14), let $\{\hat{\mathbf{p}}\}$ be the points (if any) of $\mathcal{R}$ which are collinear with $\mathbf{p}$ and the origin. Our candidate estimates $\{\hat{x}\}$ for x_{corr} are obtained applying (15) to $\{\hat{\mathbf{p}}\}$. Note that relation (15) works only for this example: in general, the mapping $\mathbf{p} \to x$ is obtained using a lookup table, as specified in section 3.

Let $\mathbf{p}$ belong to the first quadrant ($p_1 > 0, p_2 > 0$), and let $\gamma = p_2/p_1$ be the angular coefficient of the radial line passing through the origin and through $\mathbf{p}$. The radial line crosses $\mathcal{R}$ in points $\{\hat{\mathbf{p}}\}$ with

$$\hat{p}_1 = \frac{B \pm \sqrt{B^2 + (A^2 - B^2)\left(1 + \frac{\gamma^2}{a^2\omega^2}\right)}}{1 + \frac{\gamma^2}{a^2\omega^2}} \tag{17}$$

if the term under square root is not negative. It is easy to convince oneself that the positive instance of the square root is the correct choice for the match in our case. If the term under square root in (17) is negative, no match for $\mathbf{p}$ is provided by our algorithm. This circumstance is

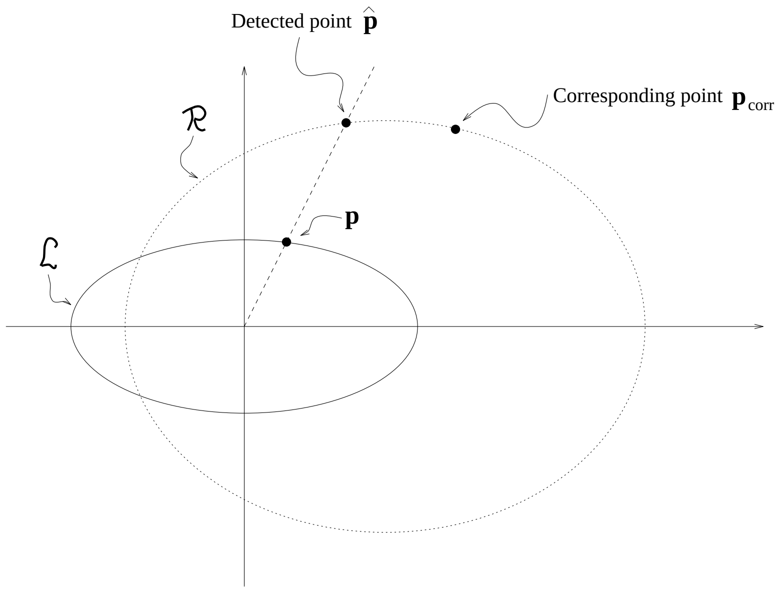


Figure 8: Intrinsic curves generated by signals $l(x)$ (solid line) and $r(x)$ (dotted line) of figure 7. The dashed line represents the radial line across $\mathbf{p}$.

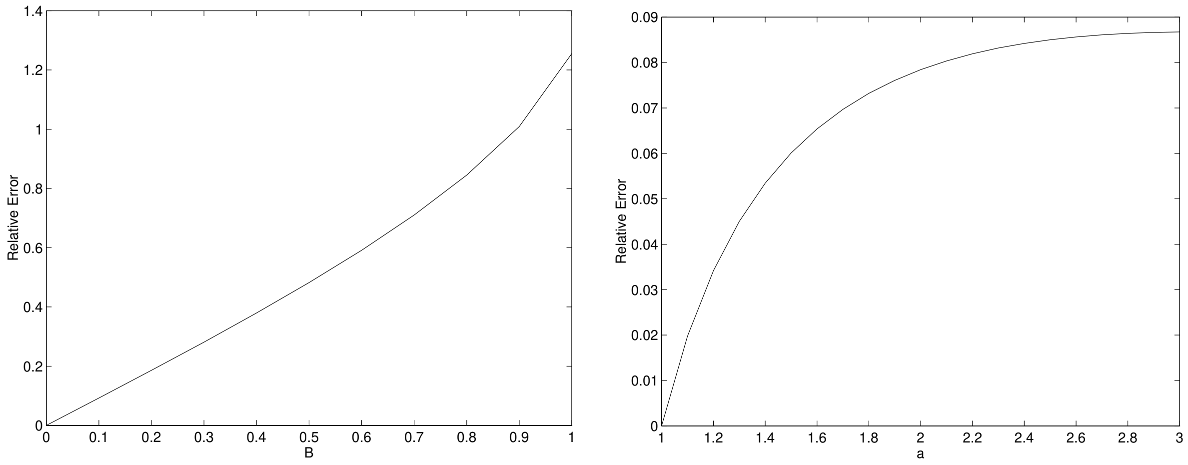
avoided if $|A| > |B|$, a condition that we will assume to be satisfied. Using (17), we see that our algorithm matches $l(x)$ at point x with $r(x)$ at point $\hat{x}$, with

$$\hat{x} = \frac{\arcsin\left(\frac{B+\sqrt{B^2+(A^2-B^2)\left(1+\frac{\gamma^2}{a^2+\omega^2}\right)}}{A\left(1+\frac{\gamma^2}{a^2+\omega^2}\right)} - \frac{B}{A}\right) - \omega d}{\omega a}. \qquad (18)$$

Figure 7 (a) shows signals $l(x)$ (solid line) and $r(x)$ (dashed line) for $A = 1.5$, $B = 0.8$, $a = 1.5$, $\omega = \pi/6$, $d = 2$. The two signals corresponds to the brightness intensity of horizontal slices of the images shown in figure 7 (b). Choosing point $\mathbf{p} = (0.253, 0.506)$ on curve $\mathcal{L}$ (corresponding to $x = 0.489$), the radial line has angular coefficient $\gamma = 2$. Curves $\mathcal{L}$ and $\mathcal{R}$ are shown in figure 8, together with the radial line. The line intersects $\mathcal{R}$ in $\hat{\mathbf{p}} = (0.582, 1.176)$, which corresponds to $\hat{x} = -1.687$. The correct match would be in $\mathbf{p}_{corr} = (1.180, 1.140)$, corresponding to $x_{corr} = -1.007$. Figure 7 (a) shows the correct match (indicated by the solid segment), and the estimated one (dashed segment).

Due to the terms $a \neq 1$ and $B \neq 0$, the match produced by our algorithm suffers from error $|\hat{x} - x_{corr}| = 0.680$ units (for comparison, the period of the sinusoid $l(x)$ is twelve units). The algorithms of Kass [Kas84] and of Jones and Malik [JM92a] (which look for a point of $\mathcal{R}$ with the minimum L_2 or L_1 distance to $\mathbf{p}$), give, in this instance, higher disparity errors, as one can verify by observing figure 8.

We can measure, for the proposed example, the dependence of the relative error $e_{rel} \stackrel{\text{def}}{=} \frac{|\hat{x}-x_{corr}|}{|x-x_{corr}|}$ on the photometric and geometric distorsion parameters. We have seen that the system is free from errors in the cases of constant displacement and even in the presence of difference of contrast ($C \neq 1$) between the two images. It remains to determine the dependence of e_{rel} on affine



(a) (b)

Figure 9: Relative match error as a function of (a) the difference of brightness B and (b) affine geometric distorsion a between the two signals of figure 7.

geometric distorsion (a) and on the difference of brightness (B). For simplicity, we only show the dependence of e_{rel} on B and on a separately. The resulting functions are shown in figure 9 (a) and (b) respectively. Note that the errors depend on the point chosen for the match; in our simple example, we again consider the match of sinusoid $l(x)$ at point $x = 0.489$.

Figure 9 (a) indicates that the relative error depends heavily on the difference of brightness B between the two images. For example, if the right image is 20% brighter than the left image, the relative match error will be approximately 20%. This result should not be surprising: our algorithm is tuned to images with different contrast (corresponding to an isotropic expansion of curve $\mathcal{R}$ with respect to $\mathcal{L}$), but is sensitive to the difference in brightness, which induces a displacement of $\mathcal{R}$. However, as noted above, the brightness offset can be effectively reduced via a suitable preprocessing (zero-mean notch filter), and the term B exists only as a residual.

The relative error resulting from geometric affine distorsion (figure 9 (b)) is more acceptable, as it exhibits saturation and is never greater than 10%. Geometric affine distorsion ($a \neq 1$) induces the expansion of the intrinsic curve $\mathcal{R}$ along the vertical direction. Vertical lines, rather than radial, would solve this problem, but at the expense of poor performance under difference of contrast ($C \neq 1$). As mentioned before, the difference of contrast between the two images seems to be the dominant source of distorsion for intrinsic curves, and therefore we stick to our choice of radial line.

3.4 Resolving Ambiguities

In general, the trajectory $\mathbf{p}(x)$ may go through a point $\mathbf{p}$ more than once. This is the case when the intrinsic curve self-intersects in $\mathbf{p}$ (see Fig. 3 (b) and, in fact, most of the examples of intrinsic curves in this paper). Then, more than one values for $\alpha(x_0)$ is feasible from our local analysis. In other words, the local representation is not able to uniquely identify the signal in the neighborhood of x_0. Hence, to correctly match the two signals at point x_0, it is necessary to take into account the contextual information available on $\alpha(x)$. A self-intersection of the intrinsic curve generated by $l(x)$ means that the local description (as given by operators P_n) of $l(x)$ is the same at more than one point x — in other words, we have to cope with local ambiguity. The density of self-intersections of the curve depends on (i) the choice of operators P_n and (ii) the number N of such operators. As a matter of fact, increasing the dimensionality of the local description should provide a richer representation (as long as the operators are independent), as noticed also in [Kas84] and [JM92a]. The intrinsic curve representation makes such a notion apparent from a topological standpoint; for example, using only two operators, the intrinsic curves lie in a plane, and self-intersections are to be expected. With three operators, the curves live in a 3-D space, where it is more likely that a path never crosses itself. A "pathological" case is when the operators P_n are shift-invariant, and $l(x)$ is periodic. In such a case, the intrinsic curve is closed, and infinite instances for $\alpha(x)$ are available. This fact reflects the inherent ambiguity in the match of periodic signals.

In addition to these "exact" forms of ambiguity, in which an intrinsic curve goes over itself precisely, approximate ambiguity may occur, in which the curve loops close to itself. Regardless of which ambiguity is present, stapling must determine a pointwise correspondence between intrinsic curves $\mathcal{L}$ and $\mathcal{R}$. Our algorithm chooses as candidate matches $\{\hat{\mathbf{p}} \in \mathcal{R}\}$ to point $\mathbf{p} \in \mathcal{L}$, the intersections of the radial line passing through $\mathbf{p}$. More than one candidate is possible, as many points of the same image may share similar local description. Among them, we have to pick the "correct" match. In addition, we would like to assign each match a confidence measure. This is an important feature of any matching algorithm which gives dense measurements. In fact, due to the noise and to local signal characteristics, we cannot expect that every point admit a reliable match. Hence, we should provide the higher-level parts of the system with confidence measures on the computed disparity field.

Resolving the ambiguity means selecting one match from $\{\hat{\mathbf{p}}\}$. This can be done only from a global standpoint: we select a suitable "path" of matches from the sets of candidate matches relative to each sampling point of $l(x)$. In the literature, ambiguity is typically resolved by imposing constraints on the disparity field, such as *uniqueness*, *ordering* (or *monotonicity* [GLY92]), and *smoothness* [MP76],[Gri85], [PMF85],[BB81],[OK85],[MN85]. Note that also other algorithms that make use of vectorial local descriptions ([Kas84], [JM92a]) need to impose constraints on the disparity field: the notion of "closeness" in the representation space is not itself sufficient for a reliable match.

The main novelty of our approach is that *disparity values never enter our procedure to solve the ambiguity.* In fact, we work only on intrinsic curves, which have lost track of space: the inverse mapping $\mathbf{p} \to x$ is determined only after the matches have been assigned.

The first constraint we impose to resolve ambiguities comes naturally from the consideration that, in ideal conditions (no photometric distorsion, compatible geometric distorsion), curves $\mathcal{L}$ and $\mathcal{R}$ are identical. Let s be the arc length parameter on $\mathcal{L}$ (see (10)), and let Δs be the length of the arc $\mathcal{C}(\mathbf{p}_1, \mathbf{p}_2)$ between two sampling points $\mathbf{p}_1, \mathbf{p}_2$ on $\mathcal{L}$. Then, we expect the length of the (oriented) arc $\mathcal{C}(\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2)$ on $\mathcal{R}$ to be "close" to Δs. Note that we still rely on the constraints of unicity and monotonicity, as parameter s is monotone with x, but we do not need any other quality of the disparity field. In other words, we simply expect that, while $\mathbf{p}$ moves along $\mathcal{L}$, the corresponding point $\hat{\mathbf{p}}$ moves similarly on $\mathcal{R}$. The important point here is that $\mathbf{p}$ and $\hat{\mathbf{p}}$ differ not because of the disparity, but because of noise and distorsions. Hence we always have to look for the "closest similar" point on $\mathcal{R}$ that satisfies this constraint on length.

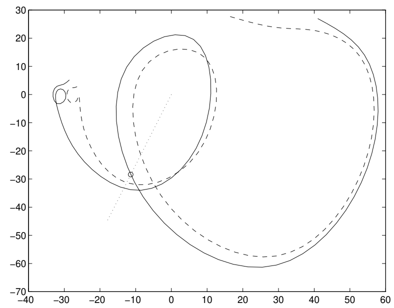


Figure 10: The correct match to $\mathbf{p}$ (circled) on the radial line is not the closest.

The previous observation suggests a first simple algorithm: assume that $\hat{\mathbf{p}}_1$ is the correct match to $\mathbf{p}_1$. Then, among the candidate matches for $\mathbf{p}_2$, choose the one such that the length of the arc connecting it to $\hat{\mathbf{p}}_1$ is closest to Δs. This procedure, however, suffers from two main shortcomings. First, we need some evidence that $\hat{\mathbf{p}}_1$ is actually a correct match. Second, we are not guaranteed that among points $\{\hat{\mathbf{p}}\}$ there is the correct match to $\mathbf{p}_2$. In fact, if $\mathbf{p}_2$ corresponds to a point which is occluded in the right image, no match to $\mathbf{p}_2$ is available. Moreover, it may happen that, due to noise, the correct match is "skipped" by the radial line. This last circumstance, although very sporadic, should be taken into account.

To circumvent these problems, we use a local measure of the quality of the match. We consider here two attributes of a candidate match $(\mathbf{p}, \hat{\mathbf{p}})$ that can be combined in a heuristic way to produce a measure of the local "goodness" $LG(\mathbf{p}, \hat{\mathbf{p}})$ of the match. One property is the Euclidean distance between $\mathbf{p}$ and $\hat{\mathbf{p}}$ (which corresponds to the distance in the representation space considered in [Kas84]). Although, as suggested by figure 10, the correct match is not necessarily the closest to $\mathbf{p}$ (see also the example of figure 8), most of the time we expect matching points to be close to each other. The second attribute is the difference between the tangents of $\mathcal{L}$ at $\mathbf{p}$ and of $\mathcal{R}$ at $\hat{\mathbf{p}}$. As noted previously, in the case of photometric distorsion ($r(x) = B + Cl(\alpha(x))$, compatible $\alpha(x)$), the tangent of the curves at corresponding points are the same. Hence, in general we expect a correct match to satisfy such a property to some extent. Practical tests have shown that this last clue is actually very powerful.

In addition to using these quality measures, we also adhere to the common practice of imposing an upper bound on the allowable displacement. In fact, intrinsic curves computed from an entire scanline are very complicated. This is because our intrinsic curves are generated by using a very simple local signal descriptor — the signal itself and its derivative. Matching these curves in their entirety would be too error prone, so scanlines must be segmented before they are matched. In our experiments we use about 30 pixels at a time, and the maximum allowed displacement must be small with respect to this width. By using richer descriptors (like in [Kas84] and [JM92a]), intrinsic curves live in spaces with more dimensions, and longer segments of the signals can be matched. To cope with large displacements, multiresolution techniques can also be exploited [Gri85],[WAH92].

In conclusion, our ambiguity resolution algorithm works on 30-pixel scanline segments, and proceeds as follows. In the first step, match candidates that are not "close and similar" enough, as measured by function LG, are eliminated. Then, the remaining match candidates, $\{\hat{\mathbf{p}}_1\}$ for $\mathbf{p}_1$, $\{\hat{\mathbf{p}}_2\}$ for $\mathbf{p}_2$, and so forth, are grouped into "coherent" segments: each segment $(\hat{\mathbf{p}}_1^i, \hat{\mathbf{p}}_2^i, \ldots)$ (where

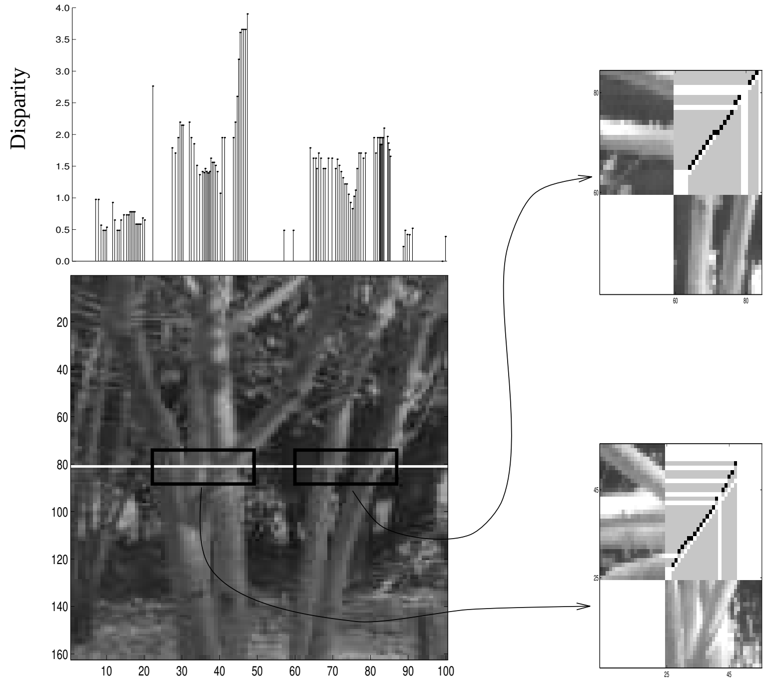


Figure 11: Sequence "Trees" (below white line: frame 1, above white line: frame 2) with the computed disparity field relative to scanline 80, and two samples of matching spaces.

$\hat{\mathbf{p}}^i_j$ is an element of $\{\hat{\mathbf{p}}_j\}$) is such that the length of the oriented arc $\mathcal{C}(\hat{\mathbf{p}}^i_j, \hat{\mathbf{p}}^i_{j+1})$ is minimum (but positive). In other words, points $(\hat{\mathbf{p}}^i_1, \hat{\mathbf{p}}^i_2, \ldots)$ follow each other closely in some arc of the oriented curve $\mathcal{R}$. A segment is terminated when (i) there are no candidate match points available to continue the ordered sequence or (ii) the length of the arc to any next match point is larger than some fixed bound (*i.e.*, we are skipping a large part of the curve).

Our final sequence of matches is obtained by selecting the "correct" segments from our list. To this purpose, we compute the average goodness (using function LG) of the points in each segment, and determine a few segment matches of high quality. Then we complete our match sequence by selecting the segments that are closer (on $\mathcal{R}$) to these reliable seed segments. This simple heuristic procedure for stapling the two intrinsic curves together has given satisfactory results in our experimental tests, summarized in figures 11 and 12.

We have chosen two couples of outdoor images, namely two successive frames of the test sequence "Trees" (figure 11, see also figure 1) from SRI, and two succeding frames of sequence "Library" (figure 12), from the movie "Wings of Desire" directed by Wim Wenders. In both cases, the camera was moving roughly horizontally (i.e., parallel to the scanlines).

We have preprocessed the images with a Gaussian filter of variance $\sigma^2 = 1$ pixel in the case of sequence "Trees" and $\sigma^2 = 3$ pixel in the case of sequence "Library". The average intensity value, computed along the whole scanline, has been subtracted from each frame to reduce brightness difference between the images to be matched (see section 3). The derivative of the signals has been computed using a simple forward–backward difference scheme.

The left part of sequence "Trees" exhibits a very articulated disparity field, induced by the sharp depth discontinuities along the boundaries of the branches of the trees. The central part

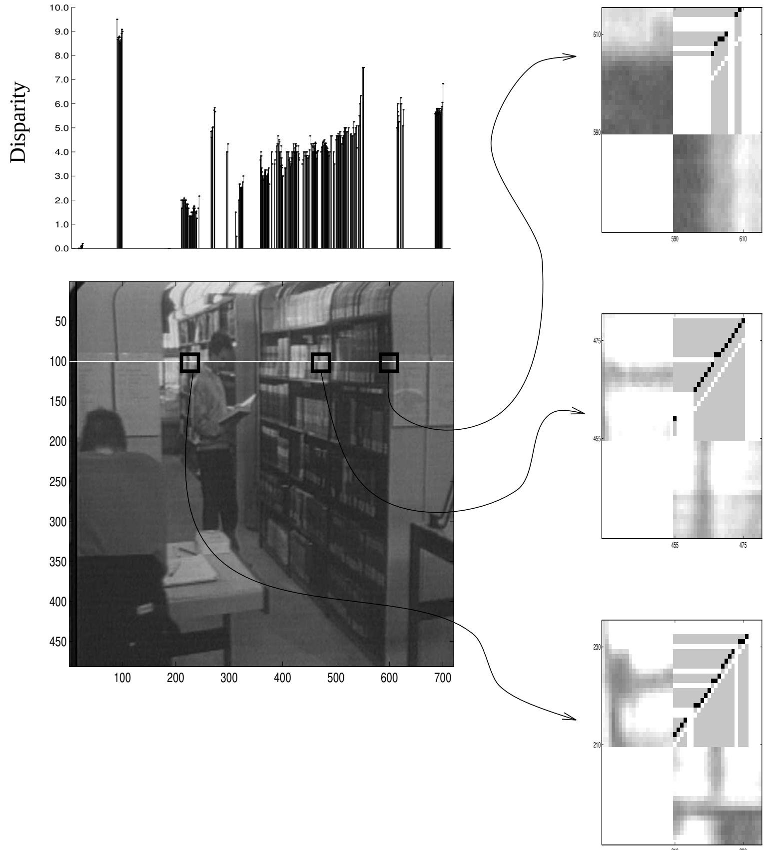


Figure 12: Sequence "Library" (below white line: frame 1, above white line: frame 2) with the computed disparity field relative to scanline 100, and three samples of matching spaces.

of figure 11 shows the corresponding image segment: the half image below the white line belongs to the first frame, the one above the white line belongs to the second frame. The disparity field corresponding to scanline 80, pixels 1–100, is shown above the image. No post–processing (e.g., filtering) has been applied to the results. Note that in some parts we have produced a dense disparity field while in other ones no measurement was available. This is due to the following reasons: (i) the intrinsic curves are sampled with uniform arclength period, which induces the nonuniform sampling period of the measurement field, and (ii) the disparity is not computed where the "goodness of match" function LG gives too small a confidence. As pointed out in section 3.2, dense measurements are characteristic of high signal busyness regions, where disparity estimates are more reliable. Figure 11 shows that the computed disparity field follows the depth discontinuities of the scene very tightly, as can be checked in the two *search planes* [OK85] (also called *matching spaces* [GLY92]) shown in the rightmost part of the figure.

Sequence "Library" (Figure 12) is characterized by a wide disparity range (from less than 0.5 pixels corresponding to the back of the room, to approximately 10 pixels at the edge of the bookshelves). The measurements are from scanline 100, pixels 1–700. Both the disparity jump corresponding to the standing person's head (pixels 260–300) and the ramp corresponding to the books on the shelf, are well tracked by our system. Note that the variance of the estimates is not negligible, as the images are quite noisy. However, the measurements are very dense, and a simple post–processing (e.g., median filtering [JM92a]) would "clean" effectively the computed disparity field.

4 Occlusions Stand Out

Any stereo algorithm must cope with occlusions. On one hand, occlusions clearly mark depth boundaries, and can be profitably exploited for object recognition, motion segmentation, and adaptive filtering. On the other hand, occlusions are important for correctly driving the match process: a stereo technique that does not take occlusions into account explicitly will suffer from serious inaccuracies and mismatches near object boundaries — the most "interesting" locations.

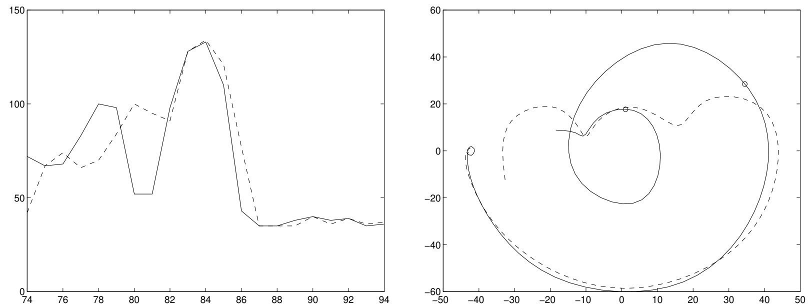
A number of researchers have dealt with the problem of occlusions in stereo, among whom we mention Toh and Forrest [TF90], Little and Gillet [LG90], Jones and Malik [JM92a], Belhumeur and Mumford [BM92], Geiger, Ladendorf and Yuille [GLY92], Intille and Bobick [IB94]. The robust and accurate detection of occlusions, however, seems still an open problem.

We will call *left occlusion* a situation where points of the left image $l(x)$ do not have counterparts in the right image $r(x)$ (see figure 13(a)) and *right occlusion* the symmetric case (see figure 14(a)).

Let us examine the effect of occlusions on the disparity field, and what problems we should expect (see also [GLY92], [IB94]). Assume that we are trying to match every point of $l(x)$ with some point of $r(x)$.

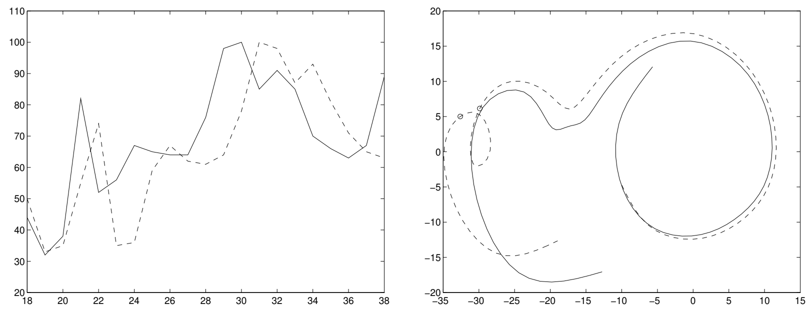
1. *Left occlusions:* the disparity is not defined here. The problem is to avoid false matches, or *false positives* [Kas84]. Ambiguity is the dangerous factor.

2. *Right occlusions:* the disparity field is discontinuous. Problems arise if we are using smoothing [MP76] or disparity gradient [PMF85] constraints on the disparity field to resolve ambiguity, as such constraints are not satisfied here.



(a) (b)

Figure 13: Scanline 95 from the image of figure 1, pixels # 74–94 (solid line: $l(x)$, dashed line: $r(x)$). (a) Intensity. The part of $l(x)$ form pixel 79 to pixel 81 is not matched by $r(x)$. (b) Intrinsic curves. The arc of $\mathcal{L}$ between the two circled points is not matched in $\mathcal{R}$.



(a) (b)

Figure 14: Scanline 92 from the image of figure 1, pixels 18–38 (solid line: $l(x)$, dashed line: $r(x)$). (a) Intensity. The part of $r(x)$ form pixel 23 to pixel 25 is not matched by $l(x)$. (b) Intrinsic curves. The arc of $\mathcal{R}$ between the two circled points is not matched in $\mathcal{L}$.

Intrinsic curves are a novel and profitable framework to detect occlusions. In fact, a left occlusion manifests itself as an arc of curve of $\mathcal{L}$ which is not matched by $\mathcal{R}$, and similarly does a right occlusion. Just before and just after the unmatched arc of curve of $\mathcal{L}$ in a left occlusion (or of $\mathcal{R}$ in a right occlusion), the curves are expected to coincide. This situation appears clearly in figures 13(b) and 13(b), where the intrinsic curves of the signals of figures 13(a) and 13(a) are depicted: occlusions stand out as "anomalous" arcs of one of the intrinsic curves.

A promising clue for detecting occlusions, currently under research, is the analysis of the topological characteristics of the two intrinsic curves. For example, both curve $\mathcal{L}$ in figure 13(b) and curve $\mathcal{R}$ in figure 14(b) contain a loop in correspondence of the occluded area, which is not matched by the other curve. This is actually a necessary condition in order to add a new segment to a curve, if the assumptions on intrinsic curves described in section 2.2 are satisfied.

In general, the presence of an unmatched loop is not by itself sufficient evidence of occlusion. Loops may be produced sometimes by noise, and we must look for a more robust topological characterization. However, it is clear that an occlusion manifests itself as a "perturbation" of only one of the two intrinsic curves in a limited region. It seems therefore that the detection and analysis of occlusions should be easier in this setting, rather than observing the profiles of the two signals in their "natural" spatial domain. In other words, the phase space is the appropriate place to look whether two signals match — or they don't.

5 From Scanlines To Images

In this paper we have made the point that search is not inherent in the stereo correspondence problem, but just in the usual way of approaching it. Our notion of intrinsic curves is a clean and useful way to think about stereo, and leads to practical matching algorithms. To the idea of associative storage and retrieval of images, intrinsic curves add the powerful constraint of connectedness. We do not propose our matching algorithm as the ultimate implementation of these concepts, but just as evidence that the approach is feasible and can deal with real, noisy images with fluctuations in brightness and contrast as well as a moderate amount of geometric distortion. Intrinsic curves are also a useful tool for the detection of occlusions, one of the most important problems in stereo matching, and lead to a natural resampling of images that emphasizes the "busier" regions where matching is more reliable.

Better algorithms can be devised, richer or more stable descriptors can be studied, occlusion detection methods can be proposed even in the single-scanline domain. We hope that the concept of compatible mappings elucidates the basic issues in the design of local image descriptors. Extensions to full images are at the same time conceptally straightforward and technically challenging, and are likely to improve performance because images have richer descriptors than scanlines. In fact, even if we restrict ourselves to the image intensities and their first derivatives, we already go from two to three descriptors. But descriptors can be made even richer through the concepts of local frequency analysis and multiresolution descriptions, both active areas of research in computer vision today.

Intrinsic curves, or surfaces, or manifolds can presumably also be *tracked* usefully, a direction of research that we are starting to investigate. Rather than looking at separate snapshots, we plan to track curves as a camera moves. Intrinsic curves, under ideal circumstances, do not move at all, so we expect tracking to be much easier than it is when done directly in the image domain.

References

[AB78] R.D. Arnold and T.O. Binford. Geometric constraints in stereo vision. In *Proc. SPIE*, volume 238, pages 281–292, San Diego, 1978.

[ADM81] J.K. Aggarwal, L.S. Davis, and W.N. Martin. Correspondence processes in dynamic scene analysis. *Proceedings of the IEEE*, 69(5):562–572, May 1981.

[Arn90] V.I. Arnold. *Ordinary Differential Equations.* The MIT Press, Cambridge, Massachusetts, 1990.

[BAR93] J. Ben-Arie and K.R. Rao. A novel approach for template matching by nonorthogonal image expansion. *IEEE Trans. Circ. Syst. Video Techn.*, 3(1):71–84, February 1993.

[BB81] H.H. Baker and T.O. Binford. Depth from edge and intensity based stereo. In *Proc. of the Seventh Int. Joint Conf. on Artif. Intell.*, pages 631–636, 1981.

[BM90] A. Blake and C. Marinos. Shape from texture: Estimation, isotropy and moments. *Artificial Intelligence*, 45:323–380, 1990.

[BM92] P.N. Belhumeur and D. Mumford. A Bayesian treatment of the stereo correspondence problem using half–occluded regions. In *Proc. of CVPR'92*, pages 506–512, Champaign, Illinois, 1992.

[CV92] M. Campani and A. Verri. Motion analysis from first–order properties of optical flow. *CVGIP: Image Understanding*, 56(1):90–107, 1992.

[GLY92] D. Geiger, B. Ladendorf, and A. Yuille. Occlusions and binocular stereo. In *Proc. of EECV'92*, pages 425–433, Santa Margherita Ligure, Italy, 1992.

[Gri85] W.E.L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Trans. Pattern Anal. Machine Intell.*, 7(1):17–34, 1985.

[HS92] R.M. Haralick and L.G. Shapiro. *Computer and Robot Vision*, chapter Image Matching. Addison–Wesley, 1992. (written by W. Förstner).

[IB94] S.S. Intille and A.F. Bobick. Disparity–space images and large occlusion stereo. In *Proc. of ECCV'94*, pages 179–186, Stockholm, Sweden, 1994.

[JM92a] D.G. Jones and J. Malik. A computational framework for determinining stereo correspondence from a set of linear spatial filters. In *Proc. of EECV'92*, pages 395–410, Santa Margherita Ligure, Italy, 1992.

[JM92b] D.G. Jones and J. Malik. Determining three–dimensional shape from orientation and spatial frequency disparities. In *Proc. of EECV'92*, pages 661–669, Santa Margherita Ligure, Italy, 1992.

[Kan84] K. Kanatani. Detection of surface orientation and motion from texture by a stereological technique. *Artificial Intelligence*, 23:213–237, 1984.

[Kas84] M.H. Kass. Computing stereo correspondence. Master's thesis, M.I.T., 1984.

[KVD76] J.J. Koenderink and A.J. Van Doorn. Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, 21:29–35, 1976.

[LG90] J.J. Little and W.E. Gillet. Direct evidence for occlusion in stereo and motion. In *Proc. of ECCV'90*, pages 336–340, Antibes, France, 1990.

[Mal89] S.G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(7):674–693, July 1989.

[Man94] R. Manmatha. A framework for recovering affine transforms using points, lines or image brightness. In *Proc. of CVPR'94*, pages 141–146, Seattle, Washington, 1994.

[MN85] G. Medioni and R. Nevatia. Segment–based stereo matching. *Computer Vision, Graphics, and Image Processing*, 31:2–18, 1985.

[MP76] D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976.

[MP90] J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America – A*, 7(5):923–932, 1990.

[OK85] Y. Ohta and T. Kanade. Stereo by intra– and inter–scanline search using dynamic programming. *IEEE Trans. Pattern Anal. Machine Intell.*, 7(2):139–154, March 1985.

[OS75] A.V. Oppenheim and R.W. Schafer. *Digital Signal Processing.* Prentice–Hall, Englewood Cliffs, New Jersey, 1975.

[PMF85] S.B. Pollard, J.E. Mayhew, and G.P. Frisby. PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449–470, 1985.

[SB92] B.J. Super and A.C. Bovik. Shape–from–texture by wavelet–based measurement of local spectral moments. In *Proc. of CVPR'92*, pages 296–301, Champaign, Illinois, 1992.

[SC94] J. Sato and R. Cipolla. Extracting the affine transformation from texture moments. In *Proc. of ECCV'94*, pages 165–172, Stockholm, Sweden, 1994.

[Str88] D.J. Struik. *Lectures on Classical Differential Geometry.* Dover, New York, New York, 1988.

[TF90] P.S. Toh and A.K. Forrest. Occlusion detection in early vision. In *Proc. of the Third Int. Conf. Comput. Vision*, pages 126–132, Osaka, Japan, 1990.

[VP87] A. Verri and T. Poggio. Against quantitative optical flow. In *Proc. of the First Int. Conf. Comput. Vision*, pages 171–180, London, 1987.

[WAH92] J. Weng, N. Ahuja, and T.S. Huang. Matching two perspective views. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(8):806–825, August 1992.

A Relation with SSD-Based Matching

In this Appendix, we show a possible applications of signal description via intrinsic curves to correlation–based matching techniques. We consider here the SSD (Sum of Squared Differences) criterion, which can be easily extended to other similarity measures (normalized SSD, correlation, etc.) [ADM81]. Given the endpoints of our measure interval $[x_0, x_1]$, we consider a continuous version of SSD, defined as

$$S(y) = \int_{x_0}^{x_1} (l(x-y) - r(x))^2 \, \mathrm{d}x \tag{19}$$

Then the disparity in $[x_0, x_1]$ is defined as the value y which minimizes $S(y)$.

Consider now the intrinsic curves $\mathcal{L}, \mathcal{R}$ generated by signals $l(x)$ and $r(x)$ as described in Section 2.2. Let $\mathbf{p}^l(x) = (p_1^l(x), p_2^l(x)) = (l(x), l'(x))$ be a generic point on $\mathcal{L}$, and $\mathbf{p}^r(x) = (p_1^r(x), p_2^r(x)) = (r(x), r'(x))$ a generic point on $\mathcal{R}$. Then it follows from (19) that

$$S(y) = \int_{x_0}^{x_1} \left(p_1^l(x-\bar{x}) - p_1^r(x)\right)^2 \, \mathrm{d}x \ . \tag{20}$$

Hence, minimizing $S(y)$ means minimizing the integral of the horizontal distance between $\mathbf{p}^l(x-y)$ and $\mathbf{p}^r(x)$. It is important to notice that, while we can visualize (19) by shifting signal $l(x)$ by quantity y, in the case of (20) the curves remain the same, and we simply change the parametrization of $\mathcal{L}$ according to y.

Relation (20) suggests a different disparity measure, which derives from a different parametrization of the curves. Let $s(x)$ be the arclength parameter referred, say, to curve $\mathcal{L}$. To account for the new parametrization, let us define $\mathbf{q}^l(s) \stackrel{\text{def}}{=} \mathbf{p}^l(x)$ and $\mathbf{q}^r(s) \stackrel{\text{def}}{=} \mathbf{p}^r(x)$, where s is related to x as described in section 3.2). Then, we can consider a new measure of similarity as

$$L(y) = \int_{x_0}^{x_1} (q_1^l(s_y) - q_2^r(s))^2 \, \mathrm{d}s \tag{21}$$

where $s_y(x) \stackrel{\text{def}}{=} s(x-y)$. We can see the convenience of using $L(y)$ instead of $S(y)$ by noting that a change of variable yields

$$L(y) = \int_{x_0}^{x_1} (l(x-y) - r(x))^2 \sqrt{(l'(x))^2 + (l''(x))^2} \, \mathrm{d}x \tag{22}$$

Relation (22) shows that measure $L(x)$ is actually a weighted sum of squared differences, the weight depending on the signal activity (expressed by term $\sqrt{(l'(x))^2 + (l''(x))^2}$). Areas of high signal activity typically are the most reliable for the match [HS92]. Note that in section 3.2 we devised in a similar fashion a dependence on the signal activity of the sampling intervals for the matches. Thus, the arclength parametrization of the intrinsic curves seems to be a profitable choice for signal matching.